University of Alberta

The Effect of Automatic Speech Recognition on Speaking Behaviors in Individuals With and Without Spinal Cord Injury

By

Jana Maureen Rieger ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of *Doctor of Philosophy*

in

Rehabilitation Science

Faculty of Rehabilitation Medicine

Edmonton, Alberta

Spring 2001

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-60339-3

Canada

University of Alberta

Library Release Form

Name of Author: *Jana Maureen Rieger*

Title of Thesis: *The Effect of Automatic Speech Recognition on Speaking Behaviors in Individuals With and Without Spinal Cord Injury*

Degree: *Doctor of Philosophy*

Year this Degree Granted: *2001*

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

2-70 Corbett Hall
Faculty of Rehabilitation Medicine
University of Alberta
Edmonton, Alberta
T6G 2G4

Date: April 12, 2001

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *The Effect of Automatic Speech Recognition on Speaking Behaviors in Individuals With and Without Spinal Cord Injury* submitted by *Jana Maureen Rieger* in partial fulfillment of the requirements for the degree of *Doctor of Philosophy* in *Rehabilitation Science*.

Paul Hagler
Paul Hagler, Ph.D.

Al Cook, Ph.D.

Megan M. Hodge
Megan Hodge, Ph.D.

Richard Jones, Ph.D.

for
Christine Sapienza, Ph.D.

Andy Greenshaw, Ph.D.

Date: April 12, 2001

DEDICATION

This thesis is dedicated to:

Mitzie Rieger
Snu P. Amell
Daisy Cameron
Casey Klimtschuk
B.G. Lockert
Bunrab and Lohng Sahks Wood
Bob and Joe Carson
Sah C. Carson
Heidi Lavertu
Ginger Frazer
Myrtle Pearlman
Pep Skinner
Mac Gallagher
Kohkoe Ouimet
Gopa Marshall

ABSTRACT

This investigation was an initial exploration of the speech-breathing and speech-production behaviors associated with the use of automatic speech recognition (ASR) software for dictation of spontaneous and scripted material by individuals with and without spinal cord injury (SCI). Surface-motion magnetometry and audiotape records were used in data collection. Speech-breathing variables included percent vital capacity inspired and expired, mean volume expired, inspiratory pause time, and duty cycle of the respiratory system for speech. Speech-production variables included syllables per breath group, frequency of breath groups, frequency of apnea, time needed for dictation and number of words spoken during a dictation task. Kinematic records were used to describe rib cage contribution to lung volume exchange, initiation and termination of speech relative to resting expiratory level, and background configuration of the chest wall. Twelve individuals participated, six with SCI and six able-bodied cohorts matched for age, sex and height. Subjects dictated with continuous-speech ASR, discrete-word ASR and no ASR. For all variables, differences amongst dictation conditions were most obvious. Trends towards differences existed between speaker groups, but no significant differences were found. Significant differences existed between speech samples for one dependent variable only (duty cycle). The differences among dictation conditions suggest that discrete-word ASR perturbs inspiratory and expiratory volume exchanges during speech from normal, while continuous-speech ASR does not. However, for duty cycle and inspiratory pause time, both ASR systems perturbed the speech-breathing system in a manner that differed from natural speech. Dictation with both discrete-word and continuous-speech ASR resulted in a decrease in the number of syllables produced per

breath group, and increases in the frequency of breath groups and apnea, with differences from normal being greater for dictation with discrete-word ASR. In addition, when participants dictated with either type of ASR, the amount of time and number of words produced were significantly greater than that associated with production of the same message without ASR, requiring "more work" on the part of the speaker and ultimately reducing the efficiency with which a message was produced. From a human factors perspective, these results suggest that ASR software, especially discrete-word ASR, has the potential to prolong speaking time and increase energy expenditure during dictation, thereby increasing speech workloads and the potential for overuse of the laryngeal system, the respiratory system, or both.

ACKNOWLEDGEMENTS

I am indebted to many people who have contributed to this research. First and foremost, I would like to thank the individuals who participated as subjects. Without your participation, this research could not have happened. As well, I would like to thank the Canadian Paraplegic Association in Edmonton for their assistance with subject recruitment.

I am also indebted to my colleagues who supported me through this process. Special mention goes to Tyler Amell, my friend and colleague within the Faculty of Rehabilitation Medicine, who supported my ideas and was an incredible sounding board throughout the whole process. I would also like to acknowledge the support and care that I received from my friends and colleagues at COMPRU (Craniofacial Osseointegration and Maxillofacial Prosthetic Rehabilitation Unit). Additionally, I would like to mention two clinicians and friends from the Glenrose Rehabilitation Hospital, Sheila Ennis and Connie Zalmanowitz, who not only supported me through the process but also participated as voice experts for certain portions of the study. As well, I extend gratitude to Amar Lallh for her participation as a voice expert. Finally, my sincere thanks go to Sharon Hundert from the University Hospital for her constant support and empathy.

I am indebted to other individuals who were instrumental in the process. First, I would like to thank Dr. Anne Rochet for her guidance throughout the course of the thesis. In addition, I would like to thank Dr. Paul Hagler for his guidance through the final course of the thesis. I would also like to thank Dr. Todd Rogers and Tess Dawber for their assistance with the statistical analyses. And finally, I would like to thank Dr. Al Cook, Dr. Megan Hodge, Dr. Richard Jones, and Dr. Andy Greenshaw from the University of Alberta who served as committee members and examiners at the thesis defense.

I would like to thank the Alberta Heritage Foundation for Medical Research for their financial support of this thesis project over the past two years, and the University of Alberta and Province of Alberta for their financial support prior to that.

Last, but definitely not least, I would like to thank my parents and family for their never-ending understanding and support throughout the past four and a half years of doctoral studies. I am so proud to be part of such an incredible group of people. In addition, I would like to thank my group of wonderful friends, you know who you are, for keeping my life filled with fun and excitement.

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION 1
BACKGROUND 1
Automatic Speech Recognition Systems 2
Current Limitations of Commercially Available ASR Systems 5
State of Knowledge 11
Normal Speech Production 12
Fatigue of the Voice 14
Theories of Speech Regulation/Control and Servomechanisms 16
ASR and the Regulation of Speech Input 19
Workload and Efficiency in the Context of ASR 22
Implications of ASR Limitations for Users with Spinal Cord Injury 24
Rationale for the Current Investigation 26
CHAPTER 2 - HYPOTHESES AND METHODS 29
PURPOSE 29
EXPERIMENTAL VARIABLES 29
Independent Variables 29
Dependent Variables 29
EXPERIMENTAL DESIGN 31
HYPOTHESES & HYPOTHETICAL OUTCOMES 31
METHOD 34
Participants 34
Procedure 45
Data Management and Analysis 53
Flowchart of the Study 57

TABLE OF CONTENTS (CONTINUED)

CHAPTER 3 - SPEECH BREATHING RESULTS **58**

Purpose 58

Variables 58

Hypotheses & Hypothetical Outcomes 58

Method 59

Data Processing and Analysis *59*

Statistical Analyses *65*

Results 66

Mean Volume (%VC) per Pre-Speech Inspiration *69*

Mean Volume (%VC) per Expiratory Segment for Speech *70*

Average Absolute Volume Expenditure per Expiratory Segment for Speech *722*

Duty Cycle of the Respiratory System within the Expiratory Phase for Speech *744*

Average Duration of Inspiratory Pause Time *788*

Measurement Reliability 80

Discussion 80

Mean Volume (%VC) per Pre-Speech Inspiration *80*

Mean Volume (%VC) per Expiratory Segment for Speech *85*

Average Absolute Volume Expenditure per Expiratory Segment for Speech *88*

Duty Cycle of the Respiratory System within the Expiratory Phase for Speech *93*

Average Duration of Inspiratory Pause Time *96*

Speech-Breathing Data Concluding Remarks *100*

CHAPTER 4 - RESPIRATORY KINEMATICS RESULTS **104**

Purpose 104

Variables 104

Method 104

Data Collection, Processing and Analysis *104*

TABLE OF CONTENTS (CONTINUED)

RESULTS AND DISCUSSION 112

Relative Chest Wall Part Contribution to Lung Volume Exchange *112*

Initiation and Termination of Speech Relative to REL *124*

Lung Volume Excursions *131*

Background Configuration of the Chest Wall *132*

Kinematic Data Concluding Remarks *153*

CHAPTER 5 - SPEECH PRODUCTION RESULTS **156**

PURPOSE 156

VARIABLES 156

HYPOTHESES & HYPOTHETICAL OUTCOMES 156

METHOD 157

Data Processing and Analysis *157*

Statistical Analyses *159*

RESULTS 160

The California Passage *160*

Dictation of a Letter *172*

Measurement Reliability *179*

DISCUSSION 179

Average Number of Syllables per Breath Group *180*

Frequency of Breath Groups *186*

Frequency of Apneic Episodes *191*

Time needed for Dictation of the California Passage *197*

Number of Words Used *198*

Speech-Production Data Concluding Remarks *201*

TABLE OF CONTENTS (CONTINUED)

CHAPTER 6 – SUMMARY AND CONCLUSIONS 204

SUMMARY OF FINDINGS 204

Regulation of Speech without ASR 204

ASR and the Regulation of Speech 205

ASR and the Speaking Mechanism 212

CLINICAL SIGNIFICANCE OF OUTCOMES 216

CRITIQUE OF THE STUDY: DESIGN VALIDITY - STRENGTHS AND LIMITATIONS 220

Internal Validity 221

External Validity 225

Statistical Conclusion Validity 228

Construct Validity of Causes and Effects 230

CONCLUSIONS 232

SUGGESTIONS FOR FUTURE STUDY 235

REFERENCES 238

APPENDICES 250

APPENDIX A 250
Information For Potential Participants 250

APPENDIX B 253
Subject's History Form: Language, Speech And Voice 253

APPENDIX C 256
Consent Template 256

APPENDIX D 257
California Passage 257

APPENDIX E 258
Dictating A Letter 258

LIST OF TABLES

TABLE 2- 1. HYPOTHETICAL OUTCOMES FOR SPEECH-BREATHING VARIABLES . 32

TABLE 2- 2. HYPOTHETICAL OUTCOMES FOR SPEECH-PRODUCTION VARIABLES. 33

TABLE 2- 3. PHYSICAL CHARACTERISTICS OF EACH PARTICIPANT. 35

TABLE 2- 4. VARIABLES FOR WHICH SKEWNESS AND KURTOSIS RATIO VALUES FELL OUTSIDE THE RANGE OF A NORMAL DISTRIBUTION. 53

TABLE 3- 1. HYPOTHETICAL OUTCOMES FOR SPEECH-BREATHING VARIABLES . 59

TABLE 3- 2. MEAN SCORES AND STANDARD DEVIATIONS OBTAINED FOR ALL THE SPEECH-BREATHING VARIABLES 67

TABLE 3- 3. SUMMARY OF ANOVA RESULTS FOR WITHIN-SUBJECTS RESULTS FOR EACH SPEECH-BREATHING VARIABLE. 68

TABLE 3- 4. SUMMARY OF ANOVA RESULTS FOR BETWEEN-GROUPS RESULTS FOR EACH SPEECH-BREATHING VARIABLE. 69

TABLE 4- 1. ISOVOLUME SLOPE AND % RC CONTRIBUTION FOR TIDAL BREATHING AND SPEAKING TASKS 112

TABLE 4- 2. LUNG-VOLUME INITIATION LEVELS FOR THE CONTROL SUBJECTS 128

TABLE 4- 3. LUNG-VOLUME INITIATION LEVELS FOR THE SUBJECTS WITH SCI . 128

TABLE 4- 4. LUNG-VOLUME TERMINATION LEVELS FOR THE CONTROL SUBJECTS 128

TABLE 4- 5. LUNG-VOLUME TERMINATION LEVELS FOR THE SUBJECTS WITH SCI . 128

TABLE 5- 1. HYPOTHETICAL OUTCOMES FOR SPEECH PRODUCTION VARIABLES BETWEEN GROUPS AND ACROSS DICTATION CONDITIONS. 157

TABLE 5- 2. MEANS AND STANDARD DEVIATIONS FOR THE SPEECH-PRODUCTION VARIABLES ANALYZED DURING DICTATION OF THE CALIFORNIA PASSAGE . 161

TABLE 5- 3. SUMMARY OF ANOVA RESULTS FOR WITHIN-SUBJECTS RESULTS FOR EACH SPEECH PRODUCTION VARIABLE DURING DICTATION OF THE CALIFORNIA PASSAGE. 162

TABLE 5- 4. SUMMARY OF ANOVA RESULTS FOR BETWEEN-GROUPS RESULTS FOR EACH SPEECH PRODUCTION VARIABLE DURING DICTATION OF THE CALIFORNIA PASSAGE. 162

TABLE 5- 5. MEANS AND STANDARD DEVIATIONS FOR THE SPEECH-PRODUCTION VARIABLES ANALYZED DURING THE FIRST 3 MINUTES OF DICTATION OF A LETTER. 172

TABLE 5- 6. SUMMARY OF ANOVA RESULTS FOR WITHIN-SUBJECTS RESULTS FOR EACH SPEECH PRODUCTION VARIABLE DURING SPONTANEOUS DICTATION OF A LETTER.. 173

TABLE 5- 7. SUMMARY OF ANOVA RESULTS FOR BETWEEN-GROUPS RESULTS FOR EACH SPEECH PRODUCTION VARIABLE DURING SPONTANEOUS DICTATION OF A LETTER. 173

LIST OF TABLES (CONTINUED)

TABLE 5- 8. SYLLABLES PER BREATH GROUP RECORDED IN THE PRESENT STUDY AND OTHERS FOR NON-DISABLED INDIVIDUALS READING ALOUD UNDER NATURAL SPEAKING CONDITIONS. 180

TABLE 5- 9. MEAN WORD-PER-MINUTE RATES FOR ALL PARTICIPANTS ACROSS DICTATION CONDITIONS FOR BOTH SPEAKING TASKS. 183

TABLE 5- 10. MEAN NUMBER OF SYLLABLES PRODUCED PER BREATH GROUP BY SPEAKERS IN BOTH GROUPS DURING DICTATION WITH NO ASR FOR BOTH SPEAKING TASKS 184

TABLE 5- 11. RATIO OF SPEECH-PRODUCTION EFFICIENCY FOR ONE UNIT OF WORK NORMALIZED TO THE CONTROL CONDITION DURING DICTATION OF THE CALIFORNIA PASSAGE 189

TABLE 5- 12. RATIO OF SPEECH-PRODUCTION EFFICIENCY FOR ONE UNIT OF WORK NORMALIZED TO THE CONTROL CONDITION DURING SPONTANEOUS DICTATION OF A LETTER. 190

TABLE 5- 13. MEAN BREATH GROUPS PER MINUTE FOR CONTROL SUBJECTS ACROSS DICTATION CONDITIONS FOR BOTH SPEECH SAMPLES. 191

TABLE 5- 14. MEAN FREQUENCY OF APNEA PER MINUTE FOR ALL PARTICIPANTS ACROSS DICTATION CONDITIONS FOR BOTH SPEAKING TASKS. 194

TABLE 5- 15. AVERAGE NUMBER OF WORDS SPOKEN, RELATED TO BOTH THE MESSAGE AND TO CORRECTIVE UTTERANCES FOR MISRECOGNITIONS, DURING DICTATION OF THE CALIFORNIA PASSAGE 200

TABLE 5- 16. AVERAGE NUMBER OF WORDS SPOKEN, RELATED TO BOTH THE MESSAGE AND TO CORRECTIVE UTTERANCES FOR MISRECOGNITIONS, DURING SPONTANEOUS DICTATION 200

TABLE 6- 1. CHANGES IN SPEECH-BREATHING AND SPEECH-PRODUCTION BEHAVIORS DURING DICTATION WITH CONTINUOUS-SPEECH ASR AS COMPARED TO DICTATION WITH NO ASR. 207

TABLE 6- 2. CHANGES IN SPEECH-BREATHING AND SPEECH-PRODUCTION BEHAVIORS DURING DICTATION WITH DISCRETE-WORD ASR AS COMPARED TO DICTATION WITH NO ASR. 208

TABLE 6- 3. CHANGES IN SPEECH-BREATHING AND SPEECH-PRODUCTION BEHAVIORS WHEN DICTATING WITH DISCRETE-WORD ASR AS COMPARED TO DICTATION WITH CONTINUOUS-SPEECH ASR. 210

TABLE 6- 4. TABLE OF SIGNIFICANT DIFFERENCES FOUND BETWEEN SPEAKER GROUPS, ACROSS DICTATION CONDITIONS, AND BETWEEN SPEECH SAMPLES. 217

LIST OF FIGURES

FIGURE 2- 1. INSTRUMENTAL ARRAY USED FOR COLLECTION OF SPEECH AND BREATHING SIGNALS. 38

FIGURE 2- 2. RELATIVE DIAMETER CHART. 42

FIGURE 2- 3. HYPOTHETICAL RELATIVE VOLUME CHART FOR A NEUROLOGICALLY NORMAL SPEAKER IN THE UPRIGHT BODY POSITION. 44

FIGURE 2- 4. FLOWCHART OUTLINING THE EXPERIMENTAL DESIGN FOR EACH SUBSET OF ANALYSES. 57

FIGURE 3- 1. A SCHEMATIC OF THE RIB CAGE (RC) AND ABDOMEN (ABD) MAGNETOMETRIC WAVEFORMS FOR 5 UTTERANCE CYCLES AS THEY APPEARED ON THE OSCILLOSCOPE SCREEN. 61

FIGURE 3- 2. AUDIO, RIB CAGE (RC), AND ABDOMEN (ABD) WAVEFORMS FOR 5 RESPIRATORY CYCLES THAT SUPPORTED SPEECH AS DISPLAYED IN CSPEECH FOR A PARTICIPANT DICTATING WITH NO ASR.. 61

FIGURE 3- 3. A SCHEMATIC OF THE RIB CAGE (RC) AND ABDOMEN (ABD) MAGNETOMETRIC WAVEFORMS FOR 5 UTTERANCE CYCLES AS THEY APPEARED ON THE OSCILLOSCOPE SCREEN. 64

FIGURE 3- 4. AUDIO, RIB CAGE AND ABDOMEN WAVEFORMS FOR 5 UTTERANCE CYCLES AS VIEWED IN CSPEECH.. 65

FIGURE 3- 5. MAIN EFFECTS FOR BOTH MEAN INSPIRATORY AND MEAN EXPIRATORY VOLUMES EXPRESSED AS PERCENT VITAL CAPACITY ACROSS DICTATION CONDITIONS. 71

FIGURE 3- 6. PARTIAL ETA-SQUARED VALUES FOR MEAN VOLUME PER PRE-SPEECH INSPIRATION. 71

FIGURE 3- 7. PARTIAL ETA-SQUARED VALUES FOR MEAN VOLUME PER EXPIRATORY SEGMENT FOR SPEECH. 72

FIGURE 3- 8. MAIN EFFECTS FOR AVERAGE ABSOLUTE VOLUME EXPENDITURE PER EXPIRATORY SEGMENT FOR SPEECH ACROSS DICTATION CONDITIONS. 73

FIGURE 3- 9. PARTIAL ETA-SQUARED VALUES FOR AVERAGE ABSOLUTE VOLUME EXPENDITURE PER EXPIRATORY SEGMENT FOR SPEECH. 74

FIGURE 3- 10. MAIN EFFECTS FOR DUTY CYCLE OF THE RESPIRATORY SYSTEM WITHIN THE EXPIRATORY PHASE FOR SPEECH ACROSS DICTATION CONDITIONS. 75

FIGURE 3- 11. PARTIAL ETA-SQUARED VALUES FOR DUTY CYCLE OF THE RESPIRATORY SYSTEM WITHIN THE EXPIRATORY PHASE FOR SPEECH. 76

FIGURE 3- 12. MAIN EFFECTS FOUND FOR DUTY CYCLE OF THE RESPIRATORY SYSTEM WITHIN THE EXPIRATORY PHASE FOR SPEECH BETWEEN SPEECH SAMPLES. 77

FIGURE 3- 13. PARTIAL ETA-SQUARED VALUES FOR DUTY CYCLE OF THE RESPIRATORY SYSTEM WITHIN THE EXPIRATORY PHASE FOR SPEECH. 77

FIGURE 3- 14. MAIN EFFECTS FOR INSPIRATORY PAUSE TIME ACROSS DICTATION CONDITIONS. 79

FIGURE 3- 15. PARTIAL ETA-SQUARED VALUES FOR INSPIRATORY PAUSE TIME. 79

LIST OF FIGURES (CONTINUED)

FIGURE 4- 1. PERCENT RIB CAGE CONTRIBUTION ACROSS DICTATION CONDITIONS FOR ALL SUBJECTS FOR DICTATION OF THE CALIFORNIA PASSAGE AND SPONTANEOUS DICTATION OF A LETTER. 120

FIGURE 4- 2. SAMPLE SPEECH-BREATHING TRACE FROM SUBJECT PM. 122

FIGURE 4- 3. KINEMATIC TRACINGS FOR A SUBJECT WITH SCI (PM) 133

FIGURE 4- 4. KINEMATIC TRACINGS FOR A SUBJECT WITH SCI (PM) 134

FIGURE 4- 5. KINEMATIC TRACINGS FOR A SUBJECT WITH SCI (PM) 135

FIGURE 4- 6. KINEMATIC TRACINGS FOR A CONTROL SUBJECT (KT) 136

FIGURE 4- 7. KINEMATIC TRACINGS FOR A SUBJECT WITH SCI (CM) 137

FIGURE 4- 8. KINEMATIC TRACINGS FOR A CONTROL SUBJECT (FM).. 138

FIGURE 4- 9. KINEMATIC TRACINGS FOR CONTROL SUBJECT (TA) 139

FIGURE 4- 10. KINEMATIC TRACINGS FOR A SUBJECT WITH SCI (JA) 140

FIGURE 4- 11. KINEMATIC TRACINGS FOR A CONTROL SUBJECT (BS) 141

FIGURE 4- 12. KINEMATIC TRACINGS FOR A SUBJECT WITH SCI (GC) 142

FIGURE 4- 13. KINEMATIC TRACINGS FOR A SUBJECT WITH SCI (GC) 143

FIGURE 4- 14. KINEMATIC TRACINGS FOR A SUBJECT WITH SCI (GC) 144

FIGURE 4- 15. KINEMATIC TRACINGS FOR A CONTROL SUBJECT (RT).. 145

FIGURE 4- 16. KINEMATIC TRACINGS FOR A SUBJECT WITH SCI (EJ) 146

FIGURE 4- 17. KINEMATIC TRACINGS FOR A CONTROL SUBJECT (GG) 147

FIGURE 5- 1. MAIN EFFECTS FOR SYLLABLES PER BREATH GROUP ACROSS DICTATION CONDITIONS DURING DICTATION OF THE CALIFORNIA PASSAGE. 163

FIGURE 5- 2. PARTIAL ETA-SQUARED VALUES FOR SYLLABLES PER BREATH GROUP. 164

FIGURE 5- 3. MAIN EFFECTS FOR FREQUENCY OF BREATH GROUPS ACROSS DICTATION CONDITIONS DURING DICTATION OF THE CALIFORNIA PASSAGE. 165

FIGURE 5- 4. PARTIAL ETA-SQUARED VALUES FOR FREQUENCY OF BREATH GROUPS. 166

FIGURE 5- 5. MAIN EFFECTS FOR FREQUENCY OF APNEA ACROSS DICTATION CONDITIONS DURING DICTATION OF THE CALIFORNIA PASSAGE. 167

FIGURE 5- 6. PARTIAL ETA-SQUARED VALUES FOR FREQUENCY OF APNEA. 167

FIGURE 5- 7. MAIN EFFECTS FOR TIME NEEDED FOR DICTATION OF THE CALIFORNIA PASSAGE ACROSS DICTATION CONDITIONS. 169

LIST OF FIGURES (CONTINUED)

FIGURE 5- 8. PARTIAL ETA-SQUARED VALUES FOR TIME NEEDED FOR DICTATION. 169

FIGURE 5- 9. MAIN EFFECTS FOR THE NUMBER OF WORDS USED IN DICTATION OF THE CALIFORNIA PASSAGE ACROSS DICTATION CONDITIONS. 171

FIGURE 5- 10. PARTIAL ETA-SQUARED VALUES FOR NUMBER OF WORDS USED IN DICTATION OF THE CALIFORNIA PASSAGE. 171

FIGURE 5- 11. MAIN EFFECTS FOR SYLLABLES PER BREATH GROUP ACROSS DICTATION CONDITIONS DURING SPONTANEOUS DICTATION OF A LETTER. 174

FIGURE 5- 12. PARTIAL ETA-SQUARED VALUES FOR SYLLABLES PER BREATH GROUP DURING DICTATION OF A LETTER. 175

FIGURE 5- 13. MAIN EFFECTS FOR FREQUENCY OF BREATH GROUPS ACROSS DICTATION CONDITIONS DURING SPONTANEOUS DICTATION OF A LETTER. 176

FIGURE 5- 14. PARTIAL ETA-SQUARED VALUES FOR FREQUENCY OF BREATH GROUPS DURING DICTATION OF A LETTER. 177

FIGURE 5- 15. MAIN EFFECTS FOR NUMBER OF WORDS SPOKEN DURING SPONTANEOUS DICTATION OF A LETTER ACROSS DICTATION CONDITIONS. 178

FIGURE 5- 16. PARTIAL ETA-SQUARED VALUES FOR NUMBER OF WORDS SPOKEN DURING SPONTANEOUS DICTATION OF A LETTER. 179

FIGURE 5- 17. TRENDS IN THE DATA FOR SYLLABLES PRODUCED PER BREATH GROUP BETWEEN SPEECH SAMPLES FOR ALL PARTICIPANTS 186

FIGURE 5- 18 A - B. FIGURE 4-18A (LEFT) REPRESENTS THE FREQUENCY OF APNEA DURING DICTATION OF THE CALIFORNIA PASSAGE FOR BOTH GROUPS. 195

CHAPTER 1 - Introduction

Background

Automatic speech recognition (ASR) technology as a user-computer interface is becoming a popular and affordable input device for word processing and other computer-based tasks. With the utilization of a microphone, this technology allows individuals to use speech to control computer functions that could once only be controlled by a keyboard and mouse. As is evidenced by their steadily increasing development budgets (Rowan, 1997), companies that develop and market this technology have been escalating their research efforts in a steadfast manner since 1995 to develop software that can more efficiently recognize speech. It is predicted that the market for ASR technology will reach $8-billion by the year 2001 (Moore, 1999). The advent of ASR software has provided a computer-input option for individuals with sensory or motor impairments such as blindness, high-level spinal cord injury or repetitive-strain injury of the upper limbs. In addition, ASR is being touted as an economically sound option for businesses that wish to decrease secretarial personnel expenses and as a convenience option for those who prefer to use dictation instead of a manual input device such as the keyboard.

Although ASR technology has evolved rapidly from the laboratory to the office and home, it has limitations that constrain the user's spoken interaction with the computer. Compensations made by the user to adjust to these limitations may be detrimental to the health of the voice. To understand the limitations of current ASR systems, it is necessary first to understand some basic information about the speech recognition process, the different types of systems that are available, and the style of speech input each of these systems requires. This information will be presented before the limitations of current speech recognition systems are discussed.

Automatic Speech Recognition Systems

Statistical pattern recognition is the basis of most speech recognition products (Strik & Cucchiarini, 1999). At its most basic level, the recognition process begins with parameter measurement of the speech signal. This is the stage at which the acoustic events of the speech signal are represented as a set of parameters. This representation is often accomplished by computing the short-term spectral envelope of the speech signal through cepstral analysis, retaining approximately the first 12 cepstral coefficients, treating these coefficients as a single vector, and assigning a sequence of vectors to a phonetic Hidden Markov Model (HMM) (Makhoul & Schwartz, 1995; Rabiner & Huang, 1993). An HMM is a statistical tool that can be used to model dynamic patterns (Sitaram & Screenivas, 1997). A common method of phoneme representation with HMMs is the 3-state model: State 1 represents the left part of the phoneme; state 2 represents the middle of the phoneme; and state 3 represents the right part of the phoneme (Makhoul & Schwartz, 1995). Thus, each HMM represents a triphone, which is a sound of phone length that has right and left transitional context properties (Ainsworth, 1997). The transitional contexts allow for consideration of allophonic variation due to coarticulation. HMMs are also created for words by combining appropriate triphone models (Ainsworth, 1997).

Each HMM is then subjected to the next stage of speech recognition, which has been described as that of pattern comparison (Makhoul & Schwartz, 1995; Rabiner & Huang, 1993). This is the stage at which the incoming triphone models are compared to a set of stored reference patterns, and search algorithms are used to determine a sequence of feature vectors with the highest probability of matching the incoming signal (Makhoul

& Schwartz, 1995). A measure of the similarity between the incoming speech signal and the stored speech patterns is made (Rabiner & Huang, 1993). In order to expedite the process, speech recognition systems will have a lexicon of words to which the parameters of the phonetic HMMs are compared. Some speech recognition systems will also have a sentence grammar from which an estimate of the probability of a word sequence can be calculated (Rabiner & Huang, 1993). A common method of specification of word sequences within a grammar is the use of a trigram, which gives the probabilities of all triplets of words in a lexicon and specifies which words are acceptable at each state of the trigram (Makhoul & Schwartz, 1995; Markowitz, 1995). State 1 of a trigram represents the leftmost word, state 2 represents the middle word, and state 3 represents the rightmost word. In the trigram model, the probability of the next word is the function of the previous 2 words (Makhoul & Schwartz, 1995). Thus, just as triphones have been successful in modeling the phone sequences in a word, word trigrams have been successful in modeling likely word sequences (Ainsworth, 1997; Lippmann, 1997; Makhoul & Schwartz, 1995).

The final stage in automatic speech recognition is the decision-making stage in which similarity scores are used to decide which stored patterns best match the incoming speech pattern. Here, probabilities for particular words and word sequences that have been established through statistical modeling of phonetic and grammatical rules are used to identify the words in the incoming speech signal.

Other aspects of automatic speech recognition systems that need mention here relate to speaker dependence and independence. Speaker-dependent systems are programs in which the HMMs are created from speech samples spoken by one person, for use by that person only. In speaker-dependent systems, the user will create a model of his/her speech in a process known as enrollment. Speaker-dependent systems are often used in situations that require a small vocabulary of words, but high levels of recognition of a particular individual's speech. For example, voice control of machine operations, where both hands are required for a task and another input modality is needed, would be a situation where speaker-dependent systems may be used. Reports of such use have been described for mobile command and control vehicles, and for the logging industry (Davis, 1996; Weinstein, 1995).

On the other hand, speaker-independent systems are programs in which the HMMs are built from speech samples of many individuals. Speaker-independent systems are essential for situations that require a large vocabulary to serve multiple users, such as for dictation (Markowitz, 1995). To increase the recognition effectiveness of speaker-independent systems, most of these programs will incorporate what is known as speaker adaptation (Makhoul & Schwartz, 1995). This is a process by which the system incrementally adjusts to the speech patterns of a particular user. Speaker adaptation can occur 'on-the-fly', that is, as a person is dictating. On-the-fly adaptation can require many hours of use before recognition performance becomes adequate (Makhoul & Schwartz, 1995). Alternatively, adaptation can also occur during a process known as rapid enrollment, which is a one-time enrollment process where the speaker dictates a known speech sample (Markowitz, 1995). This speech sample can be as short as two minutes (Makhoul & Schwartz, 1995), or as long as 2 hours (Markowitz, 1995).

The final point about speech recognition systems that must be mentioned relates to the style of speech delivery that a system has been designed to manage. Two types of

ASR systems are currently available: discrete-word and continuous-speech. Discrete-word systems require that speakers pause between each word that is spoken. This facilitates identification of word boundaries for the program (Markowitz, 1995). Continuous-speech programs do not require that speakers pause between each word, and thus allow for a more natural speaking condition. Most commercially available systems, whether discrete-word or continuous-speech, are speaker-independent systems with speaker-adaptation capabilities. Although both discrete-word and continuous-speech systems are currently very popular and available to consumers, there are limitations associated with both systems that are pertinent to the focus of this research and will be considered next.

Current Limitations of Commercially Available ASR Systems

Limitations of both discrete-word and continuous-speech ASR systems are related to the manner in which a speaker must dictate to obtain successful results. Because variation in articulation, rate, and prosody has the potential to deteriorate the performance of ASR systems (Strik & Cucchiarini, 1999), both types of software require that the speech signal be characterized by little variation. This forces individuals who use the systems to produce speech more carefully than they might when communicating with a human listener.

Variation in articulation can stem from the context in which words are spoken, the style of speaking in which one is engaged, and the natural time-varying properties of speech. With respect to context, words spoken during connected speech will differ from words that are spoken in isolation (Greenberg, 1999; Makhoul & Schwartz, 1995; Strik & Cucchiarini, 1999). When expressing a message during natural connected speech, speakers will often non-intentionally alter articulation in a manner that results in phonetic results such as deletion and reduction. For example, in connected speech, a stop-plosive consonant (i.e., a sound produced by complete airway closure behind the apposition of two speech articulators [stop portion] followed by a release of air pressure built up behind the closure [plosive release]) is often characterized by reduction or deletion of the plosive release (Crystal & House, 1988a). Or frequently, the stop is deleted altogether, especially when it occurs as the last part of a syllabic unit (Greenberg, 1999). An example of another alteration in articulation that characterizes casual connected speech is known as vowel reduction. Here, speakers will produce a vowel with the tongue in a more neutral position, which changes the resulting acoustic characteristics of that vowel (e.g. "see you" pronounced as "see ya") (Gay, 1978; Greenberg, 1999).

The style of speaking in which the speaker is engaged also can have an effect on articulation; words spoken during what has been termed 'clear speech' will be characterized by articulation that is more exaggerated than speech in a situation that does not require such effort (Helfer, 1997; Picheny, Durlach & Braida, 1989; Uchanski et al., 1996). For example, an increase in the duration of phonemes is often characteristic of clear speech. Uchanski and colleagues (1996) reported an increase in duration of short vowels by 29%, voiced plosives by 43%, unvoiced fricatives by 91% and semivowels by 103% during clear speech when compared to conversational speech. Further to this, clear production of speech often includes complete realization of word-final stops followed by schwa-like sound insertions.

Finally, the natural time-varying properties of speech production, whether during clear or conversational speech, may be troublesome for the ASR system. For instance, the length of vowels in connected speech varies markedly from one utterance to another, being influenced by linguistic factors such as prepausal lengthening (i.e., vowels preceding syntactic pauses are longer than in other locations), and the consonants that surround the vowel (Crystal & House, 1988a & 1988b). The length of vowels in connected speech tends to vary more than the length of consonants (Crystal & House, 1988a & 1988b; Rabiner & Huang, 1993). Although most speech recognizers have a built-in mechanism (such as dynamic time warping) to deal with the time-varying properties of speech, the more consistently an individual can articulate, the more easily and correctly the speech recognition system can function (Rabiner & Huang, 1993).

The variation of articulation due to context, style, and natural time-variation has the potential to lead to the alteration of acoustical information across consonants and vowels during spontaneous speech, which makes the speech recognition task difficult for an ASR system (Saraclar, Nock & Khudanpur, 2000). Imposed upon these articulatory issues is the matter of overall speaking rate, which in itself has the potential to affect the variability of pronunciation.

The overall rate of speech that is successful with each system differs. Discrete-word ASR systems require that the user leave distinct pauses between each word when dictating, thereby substantially reducing speech rates. Currently, discrete-word ASR systems recognize speech rates of only 25 to 30 words per minute (WPM). Continuous-speech ASR systems require that the speaker dictate at a somewhat slower rate than normal but without the need for distinct pauses between words. Thus, continuous-speech systems allow a speaker to dictate much more quickly than discrete-word systems and boast recognition-of-speech rates of up to 160 WPM (Dragon Systems, Inc., 2000).

Normal speech rates vary anywhere between 150 and 250 WPM (Goldman-Eisler, 1968). Therefore, even the fastest continuous-speech recognition rates fall barely within the low-end of natural speech rates. Some of the difference in WPM rates between dictated speech and natural speech might be related to the previously discussed need for more distinct articulation, and therefore slower speech, in order to improve ASR accuracy. For example, it has been found that phone deletion, a known contributor to the diminishment of ASR accuracy, increases with increased speed of speech (Fosler-Lussier & Morgan, 1999; Greenberg, 1999). In addition to this, the number of alternate forms of pronunciation of a phone increases with faster speaking rates; a factor that is likely to confuse recognition within an ASR system (Fosler-Lussier & Morgan, 1999; Greenberg, 1999). Therefore, it is likely that a slower speaking rate promotes recognition accuracy because there will be less variability associated with the speech production process and a better match of acoustic input to stored targets.

Finally, because both ASR systems require that the speech signal be fairly predictable, the natural prosodic patterns of speech also may be limited when dictating with ASR. In past research involving ASR, less attention has been paid to the prosodic patterns of speech than has been paid to the linguistic variation of articulation with respect to recognition accuracy (Saraclar et al., 2000; Strik & Cucchiarini, 1999). In spite of this, current research suggests that the suprasegmental nature of speech will be important to understand and to incorporate into the statistical model for the recognition task if automatic speech recognition is to improve (van Kuijk & Boves, 1999). Two dimensions of prosody that may affect the accuracy of ASR systems are intonation and stress patterning.

Intonation of speech is the natural rising and falling pitch contour of connected utterance that marks declarative, interrogative, and imperative messages. Intonation also lends character to a message through speaking styles such as soft, loud or angry (Rabiner & Huang, 1993), which are often influenced by psychological reactions to environmental stress. One change in intonation that has been described with the use of ASR systems is that speakers may have a tendency to use a monopitch speaking style in order to improve recognition accuracy (Kambeyanda, Singer, & Cronk, 1997). A monopitch speaking style likely varies little in the natural rising and falling of the pitch contour.

The other prosodic feature that may be affected is that of stress patterning (van Kuijk & Boves, 1999). Stress patterning is related to the emphasis that speakers place on syllables within a word, and on words within a sentence to convey meaning. Stress patterning has characteristics of speech produced both in a "clear" manner and in a normal conversational manner. As is the case with clear speech, when stress patterning dictates that a syllable be stressed within a word, the vowel in that syllable will often be lengthened (Greenberg, 1999). Likewise, when a word is emphasized or when it comes at the end of a sentence, the vocalic portions of it will often be lengthened (Crystal & House, 1988a). In addition to lengthening, a stressed vowel will have more energy in the 500Hz-4000Hz range of the spectrum and will therefore exhibit a flatter negative spectral slope than unstressed vowels (Sluijter & Van Heuven, 1996). In contrast, vowel reduction in an unstressed syllable will result in a more neutral tongue position and, therefore, acoustic characteristics that would reflect a neutral vowel (Gay, 1978). Variations in stress patterning such as the lengthening and shortening of vowels are challenging for a speech recognition system to resolve, and misrecognition of the incoming speech signal may occur.

From this review of the literature, it is evident that ASR systems have not yet reached the point where users are able to dictate in a manner that allows for the natural variation of articulation and prosody at a conversational rate of speech. Thus, speakers are forced into speech patterns that are unnaturally restricted. The limitations associated with speech recognition are shaped by the ASR system design (e.g., the statistical-modeling algorithms and the presence of a grammar) and the style of spoken input that the system will recognize (e.g., discrete-word, continuous-speech), an issue that has recently improved with the introduction of more sophisticated Hidden Markov Modeling techniques such as those used in continuous-speech systems. However, the most current continuous-speech systems, although allowing more natural speech input than discrete-word systems, still have difficulty dealing with the natural variations in speech at a conversational rate. Some of the restrictive speech patterns that may result from the inability of ASR to deal with variation have the potential to change the natural mechanics of the respiratory, laryngeal and articulatory systems. This unnatural mechanical behavior could contribute to fatigue of the speech-production mechanism and possibly lead to a communication disorder if the components of the mechanism are constrained unnaturally for prolonged periods of time. In the upcoming discussion, several aspects of these relationships between ASR and speech production will be considered. First, however, a review of what is currently known about the effect of ASR systems on the speech system will be reviewed.

State of Knowledge

Although it seems logical that ASR technology has the potential to affect speech-breathing and speech-production habits, the clinical effects of ASR technology on able- and disable-bodied users remain largely unstudied. At the present time, assertions of vocal problems associated with the frequent use of ASR technology are mostly propagated by personal communications among users and clinicians, and by reports in the popular press. These unscientific reports state that individuals who are using ASR technology following repetitive strain injury of the upper limbs are experiencing what may be the equivalent of a repetitive strain injury in the vocal mechanism; namely laryngeal fatigue, laryngeal pain, and hoarseness associated with speech as a computer input mode (Arnaut, 1995; Gooderham, 1993; Tausz, 1997b). One recent scientific survey indicates that the use of ASR has been associated with the development of moderate-to-severe voice disorders for some users (Kambeyanda, Singer, & Cronk, 1997). In this survey, the amount of time spent using the ASR system continuously without voice rest was not related to the occurrence of voice problems. However, in the clinical trial portion of the study, severe episodes of voice loss were evident after 4 hours of ASR use. Although the Kambeyanda, Singer and Cronk study lends scientific support to a potential relationship between ASR-system constraints and vocal problems, the pathophysiological bases for the development of a voice disorder remain unstudied.

A logical first step in exploring the possible link between ASR and voice disorders is to investigate the speaking behaviors employed by users of ASR technology and evaluate them against natural speech patterns. If the limited information (Kambeyanda, Singer & Cronk, 1997) relating voice disorders and the use of ASR systems is, in fact, accurate, then evaluation of speech breathing and speech production during dictation with ASR should reveal behaviors that have been associated with the occurrence of voice disorders. The focus of this research effort was to collect fundamental information upon which further studies of the relationship between ASR and the development of vocal disorders could be based. Presentation of the rationale, experimental design and procedures of the current study will be preceded by a brief review of normal speech production, consideration of fatigue of the voice, discussion of regulation/control behaviors in speech production that are pertinent, and examination of efficiency issues surrounding dictation with ASR in non-disabled and disabled users.

Normal Speech Production

The process of normal speech production begins with a relatively brief inspiration followed by an extended expiration during which pulmonary air interacts with resonating tubes and chambers in the throat, mouth, nose and cranium to produce an audible speech signal. The primary muscles responsible for inspiration to support speech are the diaphragm and the external intercostals. When inspiration of a greater volume of air is necessary (such as for long or loud utterances), accessory muscles such as the sternocleidomastoid, scalenes, subclavius, pectoralis major and minor, and the latissimus dorsi may become active (Hixon, 1987;Warren, 1988). The inspiratory muscles function to enlarge the thorax and to consequently decrease alveolar pressure in order to create a pressure gradient that will favor the inward flow of air.

In order to produce speech, alveolar pressure must exceed atmospheric pressure so that an outward flow of air will occur. At higher lung volumes, alveolar pressure is high and the outward flow of air occurs passively as a function of the elasticity of the stretched lung tissue and gravitational force. Passive expiration continues until the thorax is returned to its resting position where there is a balance of elastic forces between the lungs and chest wall (Hixon, 1987; Warren, 1988). The external intercostals muscles may be active during expiration at higher lung volumes to 'brake' the compression of the chest wall, thereby counteracting the passive forces imposed upon the respiratory system and preventing expiratory air from 'rushing out' of the lungs. When the alveolar pressure demands for speech exceed that produced by passive compression of the chest wall, active expiration must take place. The primary muscles responsible for active expiration during speech production are those of the rib cage (e.g., internal intercostals, transverses thoracis, and subcostals), as well as the muscles of the abdomen (e.g., rectus abdominus, external and internal oblique, and transverses abdominus). These muscles contract to further decrease the size of the thorax so that the necessary alveolar pressure for speech is maintained.

The outward flow of air from the lungs during expiration is the power supply for speech. For unvoiced sounds, air from the lungs passes through the glottis and is modified by downstream articulatory structures to produce speech sounds that are aperiodic in nature. For voiced sounds, air from the lungs sets adducted vocal folds into vibration, which creates a periodic sound wave that is selectively resonated and filtered within the vocal tract. Because issues related to the health of the voice have been raised in regard to dictation with ASR, voice production will be considered further.

As has been described by others (Jiang, Lin & Hanson, 2000; Hillman et al., 1989), vocal fold vibration requires pressure development beneath the adducted vocal folds (i.e., subglottal pressure) to drive them apart and then elastic restoring forces and the Bernoulli effect to bring them back together. During voice production, this cycle of vocal fold opening and closure occurs approximately 110 times per second in an average male voice, and approximately 200 times per second in an average female voice (Boone & McFarlane, 1988; Jiang, Lin & Hanson, 2000; Kent & Read, 1992). Certain pathological conditions may produce changes in the normal vibratory patterns of the vocal folds, which, in turn, may eventually affect the behavior of the respiratory system. One such condition, vocal fatigue, has been related to the use of ASR (Kambeyanda et al., 1997) and will be discussed next.

Fatigue of the Voice

Vocal fatigue is a stress-related dysfunction of the vocal apparatus (Koufman & Blalock, 1998) associated with changes in quality, loudness and pitch of the voice (Gotaas & Starr, 1993). It is thought to be related to vocal muscle exhaustion and stress relaxation in the non-muscular tissues of the vocal folds, resulting in inefficient contraction and an inability to produce normal vocal fold tonus (Neils & Yairi, 1987; Titze, 1984; Titze, 1994). As explained by Titze (1984), fatigue of the intrinsic laryngeal muscles affects the tension that is normally applied to the folds and also affects laryngeal configuration. Hence, the vocal folds of a fatigued larynx may not approximate each other fully and may appear to be bowed along their midline apposition. In addition, respiratory muscles may tire under such conditions and may not be able to provide the system with an adequate amount of expiratory pressure for efficient voice production (Titze, 1984). Symptoms of vocal fatigue may include some or all of the following: hoarseness; pitch breaks; inability to project or use a loud voice; throat and neck pain; and inflammation, swelling, and/or bowing of the vocal folds (Eustace, Stemple, & Lee, 1996; Koufman & Blalock, 1988; Morrison, 1997; Sander & Ripich, 1983; Scherer et al., 1987; Titze, 1994).

Causes of laryngeal fatigue are believed to be related to *misuse* and/or *overuse* of the voice (Koufman & Blalock, 1988; Sander & Ripich, 1983; Sapir, Atias, & Shahar, 1990; Scherer et al., 1987). *Misuse* of the voice includes abusive vocal behaviors such as: speaking with an abnormally high or low pitch, speaking with a loud voice; beginning voicing by bringing the vocal folds together abruptly; and speaking with altered breath support, such as speaking with exhalatory support from below the resting level of the respiratory system (Hixon & Putnam, 1983). *Overuse* of the voice occurs when an individual speaks for an extended period of time, and may or may not include the abusive vocal behaviors just mentioned as characteristics of misuse. Overuse often occurs as a consequence of occupations requiring substantial voice use such as classroom teaching (Gotaas & Starr, 1993; Smith, Gray, Dove, Kirchner, & Heras, 1997; Vilkman, 2000), acting or singing (Herrington-Hall et al., 1988; Kitch & Oates, 1994; Kitch, Oates & Greenwood, 1996; Koufman & Blalock, 1988; Miller & Verdolini, 1995; Sapir, 1993; Scherer et al., 1987; Vilkman, 2000), and cheerleading or aerobic instruction (Andrews & Shank, 1983; Reich, McHenry, & Keaton, 1986). Studies seeking to experimentally induce vocal fatigue in their subjects have used techniques that simulate misuse and overuse such as: speaking loudly; speaking in noise; speaking for a prolonged period of time; speaking with an abnormal pitch; or some combination of these (Gelfer, Andrews, & Schmidt, 1991; Neils & Yairi, 1987; Scherer et al., 1987; Stemple, Stanley, & Lee, 1995).

Overuse of the voice is a serious consideration when exploring the pathophysiological bases for the reported development of voice disorders associated with the use of ASR. It is useful at this juncture to explore the mechanisms of speech production in the context of regulation/control and servomechanistic theories. Further to this, consideration of ASR within these theoretical constructs may help to explain the role that this technology plays in the creation of a speaking situation that deviates from normal and that may result in overuse of the voice, and eventually in respiratory and vocal fatigue.

Theories of Speech Regulation/Control and Servomechanisms

It is well established that different biological systems of the human body, such as blood pressure and body temperature, strive to maintain some degree of homeostasis. The regulation/control theory suggests that the characteristics fundamental to the homeostasis of these systems are: (1) regulation for the purpose of stability; and (2) control mechanisms to achieve relatively steady-state conditions (Brobeck, 1965). Therefore, a regulated system is one that responds to change in order to preserve a certain degree of constancy under different conditions (Warren, 1986). Warren has hypothesized that the theory of regulation and control may be applied to the speech system because it appears to actively regulate speech airway pressures in response to feedback about critical aerodynamic and acoustic conditions.

Aerodynamically, the speech system requires a certain amount of expiratory air pressure beneath the vocal folds (i.e., subglottal air pressure) in order to drive the folds into vibration for speech production. Subglottal air pressure serves as the energy source for voice that is shaped into sonorants by structures in the supralaryngeal vocal tract. The minimum subglottal pressure required to maintain vocal fold vibration is known as *phonation threshold pressure* (Chan & Titze, 2000; Jiang, Lin & Hanson, 2000; Titze, 1993; Titze & Sundberg, 1992). Any disturbance in the necessary levels of subglottal pressure for voice production will cause the system to respond in a regulation/control manner. For example, if the vocal folds are fatigued, they may not approximate one another completely. In order to compensate for the 'leak' of air between the incompletely approximated folds, it is likely that respiratory driving force will be increased. This is not unlike what has been found in studies of the respiratory response to a leak in the oral cavity.

Structural alterations in the oral cavity which allow air to 'leak' from what should be a closed system, whether the leak is artificially induced or results from an anatomic defect, induce compensatory respiratory responses to maintain adequate oral air pressures for speech (Dalston, Warren & Smith, 1990; Laine et al., 1988; Moon, Folkins, Smith, & Luschei, 1993; Putnam, Shelton, & Kastner, 1986; Warren, 1986; Warren, Allen, & King, 1984; Warren, Hall, & Davis, 1981; Warren, Morr, Rochet & Dalston, 1989; Warren, Rochet, Dalston, & Mayo, 1992; Warren et al, 1989). The regulation/control phenomenon in this situation is likely related to the existence of mechanoreceptors in the trachea, larynx and nasopharynx for pressure in the vocal tract acting as a detection system (Putnam, Shelton & Kastner, 1986; Warren, 1986; Warren, Rochet, Dalston & Mayo, 1992), with feedback from the detection mechanism leading to a relevant change in the respiratory system to maintain the pressure levels in the vocal tract necessary for voice and speech (Moon, Folkins, Smith, & Luschei, 1993; Putnam, Shelton, & Kastner, 1986; Warren, 1986; Warren, Allen, & King, 1984; Warren, Rochet, Dalston, & Mayo, 1992). Thus, there is support for the notion that the aerodynamic input to the speech mechanism can be rationalized through the regulation/control theory.

It has been hypothesized that the speech system also acts as a servomechanism (i.e., a self-regulating machine) by using auditory feedback to regulate and control the subsequent speech output (Attanasio, 1987; Borden & Harris, 1984; Fairbanks, 1955; Fucci, Crary, Warren & Bond, 1977; Hood, 1998; Larson, 1998; Ludlow & Cikoja, 1998). Servomechanisms can either be open- or closed-looped systems. An open-loop system is not affected by external events, whereas a closed-loop system is sensitive to external events and can self-adjust to correct for errors (Hood, 1998). As stated by Hood, the speech system is a closed-loop servomechanism in which information is sent back to the control mechanism to adjust for the detection of errors in performance. There are multiple closed-loop systems acting upon the speech system, with one of them being a loop that monitors the reception of a spoken message by a listener. Dependent upon the reaction of the listener, appropriate corrections will be made to a message generated by a speaker if necessary. Experimental support of this can be found in studies that have examined the effect of segmental and suprasegmental changes that are made when an individual is speaking in order to be understood (Helfer, 1997; Picheny, Durlach & Braida, 1989; Uchanski et al., 1996). Evidence from studies of 'clear speech' has suggested that speakers will adjust their articulation to make their speech more clear by inserting extra acoustical information and by increasing the duration of sounds such as short vowels, voiced plosives, unvoiced fricatives and semivowels. Experimental evidence such as this suggests that the closed-loop speech monitoring system can detect inaccuracies in listener perception, and the speech-production system responds to that feedback by altering the manner in which sounds are produced.

The concept of such external control on the speech system is a vital concept in understanding the effect that ASR may have on the speech production process. As will be discussed next, the ASR system enters the servomechanistic loop as a 'listener' and creates a new conversational paradigm for the users of this technology, often prompting them to respond in a regulation/control manner.

ASR and the Regulation of Speech Input

In the context of regulation/control theory and closed-loop servomechanisms, an ASR system becomes a 'listener' that influences the regulation of speech. Of concern in this context, however, is its potential to induce unhealthy control responses. Most commercial ASR technology provides the speaker with immediate feedback regarding the recognition of spoken input by displaying the dictated text on the computer screen. Therefore, the speaker knows immediately if the computer has 'understood' what was dictated. If the ASR technology has recognized the words correctly, the whole system (i.e., the speaker and the computer) has been successfully regulated to achieve recognition accuracy. If a word has not been recognized, the speaker must try to repeat that word so that successful regulation may lead to recognition accuracy.

It is the manner in which the speaker responds to the computer in order to satisfy the technology's demands that is a concern. Recall that the computer requires distinct speech with little variation in speaking patterns to recognize the user's input successfully. These requirements regulate what the speaker does with the speech production mechanism to obtain successful recognition. If the computer does not recognize what has been said, the speaker may respond with repetitions of the misrecognized segment that will likely be characterized by slower, more deliberate articulation and less variable voice characteristics. If for some reason the computer does not respond successfully to these control efforts, the speaker will have to continue to formulate and execute spoken commands for correction until successful recognition is achieved. The result of this regulatory behavior is that more words will have to be produced and speaking time will be increased, both of which may result in overuse of the vocal apparatus. This has further regulatory consequences if the voice becomes fatigued.

The laryngeal muscles in a fatigued voice may become unable to sustain adequate vocal fold approximation. If the vocal folds do not approximate each other normally, there will be a 'leak' in the laryngeal valve during voice production, and a breathy voice output. The result of incomplete vocal fold approximation will be a need for compensatory expiratory air pressure beneath the folds to maintain aerodynamic parameters required to accomplish voicing at adequate loudness levels (Jiang, Lin & Hanson, 2000; Hillman et al, 1989). With respect to the regulation/control theory, this disturbance in aerodynamic events should prompt the speech system to regulate the subglottal pressure through one or some combination of the following two events. First, the system may respond through the respiratory musculature by increasing the respiratory driving power. This may eventually tire the respiratory muscles (Titze, 1994). Alternatively, the system may choose to control the leak at the level of the vocal folds. To do so, an individual would have to try to approximate the vocal folds to a greater degree to compensate for the 'leak'. Neither of these options is likely to overcome the muscular fatigue for very long, if at all, and eventually the folds will cease vibration leaving the speaker with no voice (Colton & Casper, 1990; Hillman et al., 1989). Theoretically then, even though the respiratory-laryngeal system responds in a regulation/control manner to the vocal problem induced by the abnormal speaking behaviors associated with ASR, it may be physiologically impossible for the system to restore normal pressures within the vocal tract under chronically fatigued conditions and a pattern of vocal misuse and overuse will be perpetuated.

Considering the respiratory system further, because both discrete-word and continuous-speech ASR systems encourage speech that is abnormal in juncture and rate, the possibility of inadequate and unnatural respiratory support for speech exists. With respect to juncture, the type of speech input that each ASR system requires for successful recognition will have a bearing on the depth and frequency of inspirations and expirations made in anticipation of speech production. For example, because discrete-word systems require that pauses be inserted between words, the likelihood is high that users will tend to support the disconnected speaking pattern with unnatural respiratory behaviors such as taking smaller inspirations for speech and using smaller expirations for speech. Further to this, articulation may be slowed during dictation with ASR and when speaking more slowly for the same respiratory supply, users will produce fewer words on one breath before having to re-breathe. Thus, expiratory breath groups will occur more frequently when speaking at a slower rate than when speaking at a normal rate. As the respiratory system responds in a regulatory manner to the juncture and rate requirements of ASR, as well as to the physiological requirements for voicing under such unnatural circumstances, the work imposed on this system becomes a concern.

It appears that the respiratory muscles may be susceptible to fatigue. Buchler, Magder, and Roussos (1985) studied the effect of breathing frequency on diaphragmatic blood flow in dogs and found that as breathing frequency increased, so did diaphragmatic blood flow. These authors suggested that the energy requirements of the diaphragm increased as the frequency of breathing increased, which could lead to fatigue of the respiratory system in these animals. Further to this, McCool and colleagues (1986) found that human subjects exhibited less respiratory endurance during breathing tasks in which there was high frequency of breathing. Theoretically then, the use of extremely small lung volume excursions that accumulate over the duration of the dictation task with ASR may be more strenuous for the respiratory system than when speaking without ASR, because it necessitates that the frequency of breath groups will increase over the duration of the task in order to express the same utterance at a normal speaking rate.

All of the aforementioned changes in the speech-production process associated with the regulating effects of ASR technology have the potential to result in an increase in workload and a decrease in efficiency of that process.

Workload and Efficiency in the Context of ASR

The workload and efficiency features associated with the use of ASR systems will be important to understand in order to assess the impact of this technology on non-disabled and disabled users. As with any physical exertion, there is a certain amount of work associated with speech production that is unique to every individual. An increased speech workload can be defined from a mechanical perspective as a performance characterized by measurable changes from normal in the speech-production apparatus (e.g., increases in muscle tension, increases in breathing rate, and changes in articulatory movements) (Baber, Mellor, Graham, Noyes & Tunley, 1996). Efficiency is related to the completion of a task without the waste of such resources as time and energy, and has been defined as the ratio between an externally measurable end product and the energy that was necessary to produce it (Kroemer & Grandjean, 1997). The larger the proportion of energy that goes into production of the end product, the lower the efficiency of the work done.

For a speaker using ASR technology, the message that is dictated is, in essence, the end product. Therefore, the efficiency of the process is related to the message that is produced and amount of speaking effort that went into producing it. With respect to the regulatory influence of ASR on the speech production apparatus, a workload induced by the use of speech-recognition technology may result in changes in the expiratory air supply devoted to speech, the amount of speech produced on the available air supply, the frequency of respiratory cycles for speech, and the number of words that need to be articulated in order to generate a message. Changes in these speech-breathing and speech-production behaviors with the use of ASR have the potential to create a higher workload than that associated with normal speech production, ultimately resulting in low efficiency.

Decreases in the amount of speech produced on an expiration, increases in the number of respiratory cycles needed to support speech, and increases in the number of words needed to generate a message could be predicted speech-breathing and speech-production responses during dictation with ASR. For example, it has been established by others that WPM rates are slower when dictating with ASR. Thus, it can be expected that when speaking more slowly, users will produce fewer words on the same expiratory supply. To dictate a passage of a specified length, then, more respiratory cycles will be needed when a speaker is talking more slowly. More respiratory cycles will require more respiratory muscular energy. In addition, the production of speech not related to the message, but to commands for correction of computer misrecognitions, will inflate the number of words that must be produced to dictate that message. Therefore, efficiency of dictation of a message with ASR will be lower because the proportion of energy input for the task is higher than during dictation of the same message without ASR. This is a simple example, but is exemplary of the reciprocal relationship between workload and efficiency. It is possible that the regulating effect of ASR has the potential to impose this relationship on several speech-breathing and speech-production processes. It is the intent of this study to explore several of these processes during dictation with ASR technology in both non-disabled and disabled users.

Implications of ASR Limitations for Users with Spinal Cord Injury

Individuals with cervical spinal cord injuries have constraints imposed not only on the respiratory system, but also on the laryngeal system for speech production. Constraints on the respiratory system of individuals with quadriplegia include substantially altered breathing patterns and susceptibility to respiratory fatigue (Hoit, Banzett, Brown & Loring, 1990; Hopman et al., 1997; Loveridge, Sanii, & Dubo, 1992; Manning, McCool, Scharf, Garshick & Brown, 1992; Sinderby, et al., 1996) due to loss of voluntary control of large groups of muscles of respiration. It is possible that individuals with quadriplegia will have some inspiratory muscle function, albeit impaired, if the spinal lesion spares input to some or all of the phrenic motor nerves in the cervical spinal cord that innervate the diaphragm. Individuals with cervical spinal cord injury likely also have severely compromised expiratory function due to loss of executive motor input to the thoracic and lumbar levels of the spinal cord that serve the muscles of the rib cage and abdomen (De Troyer & Estenne, 1991; Fujiwara, Hara, Chino, 1999; Sivak, Shefner & Saxton, 1999). Thus, inspiratory and expiratory gestures for speech in individuals with cervical spinal cord injury will be limited, more or less, by the level of their injuries and extent of their paresis or paralysis.

Altered inspiratory gestures for speech that have been documented through kinematic analysis in individuals with cervical SCI include slow and shallow inspirations (Hixon & Putnam, 1983; Hoit et al., 1990). Affected aspects of expiratory gestures include the use of short breath groups when speaking and volume compression limitations that result in loudness restrictions (Hixon & Putnam, 1983; Hoit et al., 1990). Because these constraints restrict the volume of air available for speech, individuals with cervical spinal cord injury may be limited in the amount of speech that they produce on each expiratory breath thereby increasing the frequency of respiratory pump cycles they produce in order to convey a message. An increased frequency of breathing to convey a message could compound respiratory muscle weakness and lead to a premature onset of fatigue in situations where the inspiratory muscles are faced with a greater workload. Thus, in order to reduce respiratory workload, individuals with cervical SCI may employ strategies to take advantage of their available expiratory supply for speech to as great an extent as possible. Such strategies may include using the larynx more strenuously than it functions for voice production under normal circumstances to control the expiratory flow of air during speech. For example, it has been suggested that breath holding may be used as a compensatory strategy to regulate expiratory air flow in those with an impaired respiratory system secondary to paralysis of the rib cage, diaphragm and abdomen (Hixon, Putnam & Sharp, 1983).

Finally, paresis of the respiratory muscles has the potential to affect inspiratory pause times. The loss of abdominal muscle function would impair a speaker's ability to stabilize the abdominal wall during inspiration thereby decreasing the mechanical advantage for contraction of the diaphragm and resulting in longer inspiratory pause times. This, along with the time associated with each respiratory pump cycle may increase the total time that an individual with SCI spends in dictating a specific message.

Altered speech-breathing and speech-production behaviors that result from a spinal cord injury may make individuals with cervical SCI especially vulnerable to the regulatory effects of ASR. Already at a disadvantage, these speakers must deal with further changes in speech-breathing and speech-production behaviors that may result from the use of ASR. These changes may impose a further increase in the workload and decrease in the efficiency of speech-production for individuals with SCI.

Rationale for the Current Investigation

It is evident that both discrete- and continuous-ASR systems may encourage speech-breathing and speech-production behaviors that are unusual, if not unnatural. Because the speech and breathing behaviors associated with either type of ASR system have not been studied previously, it is desirable to investigate these so that the extent to which they may differ from natural speech behaviors can be described. The speech-breathing and speech-production patterns exhibited and the efficiency with which a linguistic message is conveyed can be explored to gain an understanding of the workload imposed upon a speaker by adding an ASR system to the speaking paradigm.

It is the intent of this body of work to explore the regulating effect of ASR technology on several different aspects of speech-breathing and speech-production behaviors in individuals with and without SCI. The unnatural juncture and rate of speech required during dictation with ASR have the potential to create inadequate and unnatural respiratory control for speech and will have a bearing on speech-breathing behaviors such as the depth and frequency of inspirations and expirations made in anticipation of speech production, as well as the degree to which the expiratory portions of those respiratory cycles are used for speech. Speech-production behaviors will be consequently affected by changes in speech-breathing behavior. During dictation with ASR, possible restrictions in the depth of inspirations and expirations along with a decrease in speaking rate may give rise to the production of fewer words on one breath before having to re-breathe resulting in an increase in the frequency of expiratory breath groups.

In addition to the juncture and rate issues related to dictation with ASR and their effect on speech-breathing is the issue of misrecognitions of a speaker's message and its effect on speech production. Misrecognitions will result in the need for a speaker to dictate corrective commands to the ASR technology, which will inflate the number of words spoken to convey an intended message and will result in greater speaking time. Exploration of such changes in speech breathing and speech production will lay the foundation upon which further studies, such as those that explore the characteristics of the voice during dictation, can be developed in this research frontier.

Finally, because individuals with a cervical spinal cord lesion are already predisposed to laryngeal and respiratory fatigue due to their injury (Hixon & Putnam, 1983; Hoit, Banzett, Brown & Loring, 1990; Hopman et al., 1997; Loveridge, Sanii, &

Dubo, 1992; Manning, McCool, Scharf, Garshick & Brown, 1992; Sinderby, et al., 1996), it is valuable to investigate if the demands of ASR technology lend additional strain to an already strained system and create a situation where the risk for damage to the vocal folds may be increased. Individuals with cervical SCI are the most likely to need and depend on the technology because of the physical limitations of their arms and hands. Thus, an investigation of the regulating effect of ASR technology on the speech-breathing and speech-production behaviors in this population and in non-disabled individuals is worthy of consideration.

CHAPTER 2 - Hypotheses and Methods

Purpose

The purposes of this investigation were to describe, from several different perspectives, the characteristics of speech and breathing patterns associated with spontaneous speech and reading aloud in three dictation conditions: with discrete-word ASR, with continuous-speech ASR, and with no ASR. In addition, it was also desirable to determine how speech and breathing characteristics differed among those three dictation conditions within and between two groups of users: those with and without cervical spinal cord injury.

Experimental Variables

Independent Variables

The independent variables in this study included *speech sample* (spontaneous speech & reading); *dictation condition* (no ASR, discrete-word ASR & continuous-speech ASR); and *speaker group* (spinal cord injured [SCI] & non-spinal cord injured [non-SCI]).

Dependent Variables

There were three categories of dependent variables (speech-breathing variables, speech-production variables, and kinematic variables) for which data were collected based on participants' read and spontaneous speech in all three dictation conditions. They were as follows:

1. Speech-Breathing Variables:
   (a) Mean volume (%VC) per pre-speech inspiration
   (b) Mean volume (%VC) per expiratory segment for speech

    (c) Average absolute volume expenditure per expiratory segment for speech (ml)

    (d) Duty cycle of the respiratory system within the expiratory phase for speech (% speech)

    (e) Average duration of inspiratory pause time within a given passage (ms)

2. Speech-Production Variables:

    (a) Average number of syllables per breath group (#)

    (b) Frequency of breath groups per paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)

    (c) Frequency of apneic episodes per paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)

    (d) Time needed for dictation of a paragraph read aloud (California Passage) (minutes)

    (e) Number of words used in dictation of a paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)

3. Kinematic Variables:

    (a) Rib cage contribution to lung volume exchange during dictation of a paragraph read aloud and time-limited passage of spontaneous speech (%)

    (b) Initiation and termination of speech relative to REL, and range of lung volume excursions

    (c) Background configuration of the chest wall during dictation of a paragraph read aloud and time-limited passage of spontaneous speech

Operational definitions of concepts that are central to the understanding of the dependent variables are provided below:

> *inspiratory volume (%VC)*: the lung volume relative to vital capacity that comprises the inhalatory portion of a respiratory cycle upon which speech is produced
>
> *expiratory volume (%VC)*: the lung volume relative to vital capacity that comprises the exhalatory portion of a respiratory cycle upon which speech is produced

inspiratory pause time: the duration of a pre-speech inspiratory gesture between expiratory segments for speech within a related series of utterances

duty cycle of the respiratory system within the expiratory phase for speech: the ratio of the duration of speech in the expiratory phase to the duration of the total expiratory phase

apnea: a moment in the respiratory cycle in which exhalation ceases and no volume is exchanged (i.e., a moment when the breath is "held").

breath group: the speech produced on one exhalation

Experimental Design

This study used several factorial comparisons to explore differences between and within groups and speech samples, and among dictation conditions. The speech-breathing variables were investigated via a counterbalanced three-way design with one between-subjects factor (group) and two within-subjects factors (dictation condition & speech sample). The speech-production variables were investigated via a counterbalanced two-way design with one between-subjects factor (group) and one within-subjects factor (dictation condition). The kinematic variables were described with respect to speaker groups, dictation conditions, and speech samples.

Hypotheses & Hypothetical Outcomes

Directional hypotheses for each dependent measure were developed for the independent variables of speech sample (reading aloud vs. spontaneous speech), speaker group (SCI vs. non-SCI), and dictation condition (no ASR vs. discrete-word ASR vs. continuous-speech ASR), where applicable. Because speech produced in the no-ASR condition was not influenced by the computer-user interaction associated with the other two conditions, it was considered a control condition. In addition, the individuals not affected by spinal cord injury functioned as a control group.

The hypotheses were developed based on a review of pertinent literature, on *a priori* assumptions about the constraints imposed by ASR technology on natural speech breathing and speech behaviors, and on the results of a preliminary investigation of this protocol in two persons, one with cervical SCI and an age-matched control. Hypotheses developed for all dependent variables are described in the tables that follow.

Dependent Variable	Speaker Group	Dictation Condition	Speech Sample
Mean **volume** (%VC) per pre-speech inspiration across a given passage…	will be higher for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be lowest in the discrete-ASR condition and highest in the no-ASR condition for all speakers and speech samples	will be higher for reading than for spontaneous speech for all speakers and dictation conditions
Mean **volume** (%VC) per expiratory segment for speech across a given passage…	will be higher for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be lowest in the discrete-ASR condition and highest in the no-ASR condition for all speakers and speech samples	will be higher for reading than for spontaneous speech for all speakers and dictation conditions
Average absolute **volume expenditure** per expiratory segment for speech (ml) across a given passage…	will be smaller for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be smallest when using discrete-word ASR and largest in the no-ASR dictation condition for all speakers and speech samples	will be the same for both reading and spontaneous speech for all speakers and dictation conditions
Duty cycle of the respiratory system within the expiratory phase for speech (% speech)…	will be higher for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be lowest when using discrete-word ASR and highest in the no-ASR dictation condition for all speakers and speech samples	will be higher for reading than for spontaneous speech for all speakers and dictation conditions
Average duration of **inspiratory pause time** (ms) within a given passage…	will be longer for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be shortest when using discrete-word ASR and longest in the no-ASR dictation condition for all speakers and speech samples	will be shorter for reading than for spontaneous speech for all speakers and dictation conditions

Table 2- 1. Hypothetical outcomes for speech-breathing variables between speaker groups, across dictation conditions, and between speech samples.

Dependent Variable	Speaker Group	Dictation Condition
Average **number of syllables** per breath group (#)…	will be lower for individuals with SCI than for individuals without SCI across dictation conditions	will be lowest when dictating with discrete-word ASR and highest in the no-ASR dictation condition for all speakers
Frequency of breath groups per paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)…	will be higher for individuals with SCI than for those without SCI across dictation conditions	will be highest for discrete-word ASR and lowest in the no-ASR dictation condition for all speakers
Frequency of **apneic or breath-holding episodes** per paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)…	will be higher for individuals with SCI than for those without SCI across dictation conditions	will be highest during dictation with discrete-word ASR and least in the no-ASR dictation condition for all speakers
Time needed for dictation of a paragraph read aloud (California Passage) (minutes)…	will be longer for individuals with SCI than for those without SCI across dictation conditions	will be longest for discrete-word ASR and shortest in the no-ASR dictation condition for all speakers
Number of words used in dictation of a paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)…	will not differ between individuals with SCI and those without SCI across dictation conditions	will be highest for dictation of the California Passage in the discrete-word ASR condition and lowest in the no-ASR dictation condition for all speakers across dictation conditions; the opposite will be true for the letter

Table 2- 2. Hypothetical outcomes for speech-production variables between speaker groups and across dictation conditions.

The hypotheses for the speech-breathing and speech-production variables across speaker group were based on previous research that has documented differences between individuals with and without SCI in natural speaking conditions; the control and capacity limitations imposed upon the respiratory system by a spinal cord injury have the potential to affect speech-breathing and speech-production patterns (Hixon & Hoit, 1987; Hixon & Putnam, 1983; Hoit, 1990; Hoit et al., 1990; Manning et al., 1992). The hypotheses for differences across dictation conditions were based on the expectations about the effects of ASR software on speech-breathing behaviors. That is, discrete-word ASR was expected to distort natural patterns the most, while continuous-speech ASR was expected to have intermediate effects. Hypothesized differences in speech breathing between speech samples were based on previous research that revealed differences between reading and conversational speech due to the enhanced preplanning that reading affords the respiratory system (Hodge & Rochet, 1989; Winkworth, Davis, Adams & Ellis, 1995; Winkworth, Davis, Ellis & Adams, 1994).

Method

Participants

Twelve non-treatment-seeking participants between the ages of 26 and 54 years (mean = 37) participated in this study. Table 2-3 provides an overview of participant characteristics. Six participants sustained a cervical spinal cord injury (between C-4 and C7) at least 1 year prior to inclusion in the study. This time limit was imposed as breathing patterns have been shown to adjust and stabilize in the time between injury and 1 year post-onset (Almenoff et al., 1995; Loveridge, Sanii & Dubo, 1992). The remaining 6 participants were matched for age (+/- 5 years), height (+/- 10cm), and sex to those with spinal cord injury. All participants spoke English as their native language. Only 2 participants (GC, MH) reported using speech recognition systems prior to inclusion in the study. GC reported using DragonDictate (Version 3.0, Dragon Systems, Inc., Newton, MA) with a frequency of less than 5 hours per week. MH reported using NaturallySpeaking (Version 3.02, Dragon Systems, Inc., Newton, MA) with a frequency of 5 hours per week.

Subject	Age (yrs)	Height (cm)	Sex	Level of Lesion	Time post-onset (years)	Vital capacity (litres)
PM	36	167	F	C5-6	22	2.3
KT	33	158	F	N/A	N/A	3.7
CM	31	173	F	C6-7	6	2.3
FM	26	178	F	N/A	N/A	3.5
MH	33	178	M	C5-6	16	2.0
TA	28	178	M	N/A	N/A	4.3
JA	50	175	M	C5-6	28	2.3
BS	54	173	M	N/A	N/A	3.3
GC	37	183	M	C4-5	20	1.3
RT	35	183	M	N/A	N/A	4.4
EJ	39	180	M	C6-7*	2	3.2
GG	38	183	M	N/A	N/A	5.8

Table 2- 3. Physical characteristics of each participant, level of lesion, time post-onset, and vital capacity. Individuals with SCI are paired with their age-, height- and sex-matched control. * This lesion was incomplete. N/A = not applicable.

Prior to inclusion in the study, participants were judged by the primary investigator to have speech and voice characteristics within normal limits. Expectations for neurologically normal speakers with respect to voice characteristics required that their voices were rated as "1" on the voice profile established by Wilson (1990). The acceptable range of tolerance for individuals with spinal cord injury on the Wilson Scale (1990) allowed for judgements of +2 on the laryngeal cavity tension scale. This was thought necessary due to documentation that suggests these speakers may need to use the larynx as a compensatory mechanism to control expiratory airflow, normally checked by respiratory musculature (Hixon, Putnam, & Sharp, 1983). Thus, a habitually tense voice quality may be evident in many individuals with spinal cord injury.

Individuals with pacemakers or other electronic implants were not allowed to participate due to the potential for interference between these devices and the magnetometers used to sense chest wall displacements (Astridge et al., 1993; Bourke, 1996). Histories of laryngeal, velopharyngeal, craniofacial and neurological disorders other than the spinal cord injury were criteria for exclusion due to the effect that they may have on the production of normal speech. All participants were required to be free of upper respiratory infections during all of the experimental sessions because of the possible adverse effects such infections might have on measurements of air volumes exchanged and voice quality. In addition, all participants were non-smokers as it has been demonstrated that chronic cigarette use can adversely affect pulmonary function, especially in subjects with underlying respiratory compromise (Almenoff, Spungen, Lesser & Bauman, 1995), and acoustic characteristics of the voice (Murphy & Doyle, 1987; Sorenson & Horii, 1982).

All participants passed a pure tone hearing screening at 30 dB HL under earphones for octave frequencies 250-8000 Hz in the better ear. The screening procedure was performed by the primary investigator using a non-portable audiometer (calibrated to ASHA-ISO standards) in a sound booth. This requirement was necessary due to the effect that hearing loss can have on speech breathing behaviors (Forner & Hixon, 1977) and the pitch, quality, and intensity of the voice (Higgins, Carney, & Schulte, 1994). Due to the need for visual interaction with the computer screen for ASR training and dictation feedback, all participants also were required to pass a vision screening at a distance equivalence of at least 20/40. Screenings were completed with the ScyMed VISION-card screening tool (1998, Rodram Corp., La Jolla, CA). The eye chart was mounted on the wall and participants stood or sat 30 centimeters away, as per the instructions provided by Jobe (1976). The vision screening examined visual abilities of both eyes simultaneously.

ASR Instrumentation

Two automatic speech recognition software programs were used: (a) discrete-word recognition and (b) continuous-speech recognition. The discrete-word recognition program was DragonDictate (Version 3.0, Dragon Systems, Inc., Newton, MA). The continuous-speech recognition program was Dragon NaturallySpeaking Deluxe Edition (Version 3.52, Dragon Systems, Inc., Newton, MA). Both programs were run on a personal computer with a Pentium II 300MHz processor. Participants' dictation signals were transduced by an active noise cancellation head-mounted microphone (Gentex Noise Cancelling, VXI Corp., Rollinsford, NH) provided with the ASR systems. These speech signals were digitized at 22 kHz and quantized at 16 bits by an indwelling sound card (Sound Blaster Awe 64, Creative Labs Incorporated, 1997).

Dragon Systems' continuous-speech ASR technology was selected for the following reasons: (1) the program boasts a large active vocabulary which ultimately may lead to better recognition; (2) NaturallySpeaking allows mouse control by voice which is an extremely important option for individuals who have limited or no use of their upper limbs; (3) error correction is simple to learn and use; and (4) reports indicate that accuracy of recognition is high (Padilla, 1998). Because the continuous-speech ASR engine chosen for this study was a Dragon Systems' product, DragonDictate discrete-word ASR technology, which is also a Dragon Systems' product, was selected for use to increase the concurrent validity of the intended comparisons between the discrete-word and continuous-speech systems.

Instrumentation for Speech Signal Recording

The instrumental array used to obtain speech and breathing data is shown in Figure 2-1. Two microphones were used to collect participants' dictation signals: one for the ASR software, and one for acoustic analyses and listener judgments that will form the bases for additional study beyond the scope of this report on speech and speech-breathing variables.

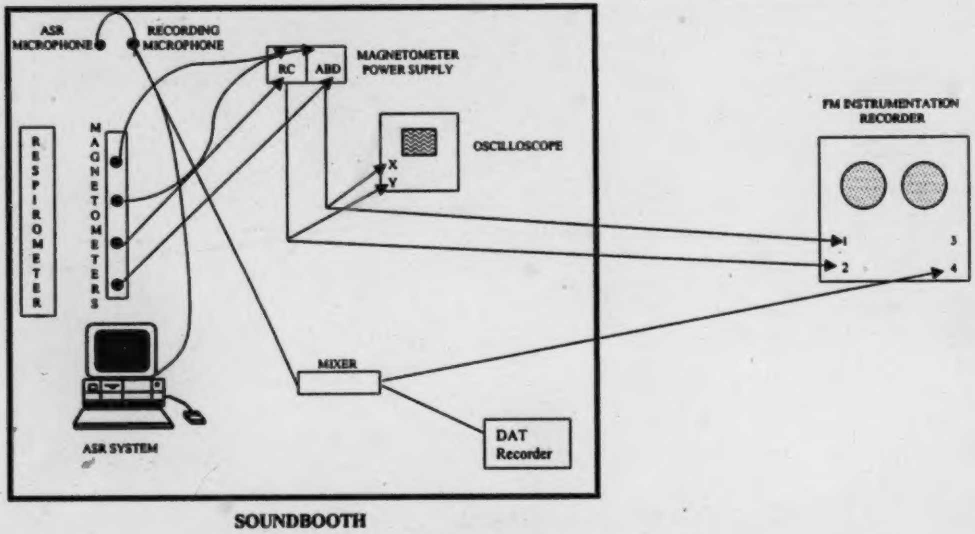
ASR MICROPHONE
RECORDING MICROPHONE
RC
ABD
MAGNETOMETER POWER SUPPLY
RESPIROMETER
MAGNETOMETERS
OSCILLOSCOPE
X
Y
FM INSTRUMENTATION RECORDER
1
2
3
4
ASR SYSTEM
MIXER
DAT Recorder
SOUNDBOOTH

Figure 2- 1. Instrumental array used for collection of speech and breathing signals.

The ASR microphone specifications indicated that its frequency response curve was relatively flat over the frequencies from 20 to 500 Hz, declined in relative output through the frequencies 500 to 5000 Hz, and dropped markedly at approximately 7000 Hz. This was not considered adequate for the simultaneous high-fidelity recordings obtained for future acoustic analyses and listener judgments. Thus, a second microphone (Shure, model SM10A, Evanston, IL) was mounted on the opposite side of the headband used to support the ASR microphone. This instrument exhibited a flat frequency response across the frequencies 30 to 15000 Hz with a dynamic range of 85 dB. Speech signals from this microphone were amplified by an audio mixer (Shure, model FP42, Evanston, IL) and digitized at 44KHz with 16-bit resolution on a digital audio tape (DAT) recorder (Tascam, DA-30, Teac Corp., Mississauga, Ontario). Intensity of speech input to the DAT recorder was controlled by adjusting the deflection of the needle on the VU meter of the Tascam mixer so that it remained between -1 and -3 on the VU scale and the LED indicators on the recorder varied accordingly. These adjustments in gain made optimal use of the dynamics of the DAT recorder without overmodulation of the speech signal. Speech signals transduced by the Shure microphone were also recorded on the 4th channel of an FM recorder (Hewlett-Packard, 3964A, San Diego, CA). During collection of preliminary breathing data from participants prior to their dictation performance, the primary investigator wore the Shure microphone so that a running commentary of the procedures could be recorded on the FM tape. During subjects' dictation, this microphone was snapped into place on the headset bearing the ASR microphone. Hereinafter, this Shure microphone will be referred to as the DAT-FM microphone to acknowledge its use with both those recording devices by both the primary investigator and the participants at different moments in the experimental procedure.

Instrumentation for Speech-Breathing Signal Recording

Surface-motion magnetometry was used to collect data regarding lung volume excursions and chest wall kinematic behaviors during the respiratory maneuvers and speech tasks included as part of the experimental protocol in this study. Measurement of chest wall diameter changes via magnetometers allows information to be obtained regarding lung volume displacement in a noninvasive manner and without the need for instrumentation at the airway opening that interferes with the speech signal and its production, or with normal respiratory behaviors. The use of magnetometry for measurement of respiratory events is based on the kinematic method first described by Konno and Mead (1967) and later refined by Hixon, Goldman, and Mead (1973). The theory behind the kinematic method is based on the premise that the chest wall is composed of two separate parts: the rib cage and the diaphragm-abdomen (hereinafter referred to as "abdomen"). The two parts function in parallel or in series depending upon whether the respiratory system is open or closed to atmosphere. In the situation where the respiratory system is open to atmosphere, such as when volume is exchanged while speaking or breathing, each part moves in parallel with and displaces volume relatively independently of the other part (Hixon, Goldman, & Mead, 1973). The total lung volume change that occurs when the respiratory system is open reflects the combined displacements of the rib cage and abdomen. Thus, data obtained from movements of the chest wall when the airway is open can be interpreted for information regarding the volume of air exchanged during breathing and speech. When the respiratory system is closed, such as in the case of breath-holding, no volume is exchanged and the two parts must move in a highly interdependent and serial fashion when required to do so. Thus, in a closed system, the volume exchanged by either the rib cage or abdomen must be equal and opposite to the other part (Hixon, Goldman, & Mead, 1973). Data obtained from movements of the chest wall in a closed system, known as isovolume maneuvers, are used for calibration purposes with surface-motion magnetometers and reveal functional relationships between the rib cage and abdomen that allow the derivation of volume-motion relationships for each part (Hixon, Goldman, & Mead, 1973).

Magnetometers have been used extensively to measure different aspects of breathing for speech and life. Magnetometers appropriate for use on children and adults consist of two electromagnetic coil pairs (each 4 cm in diameter) that are applied to the surface of the chest wall and used to transduce anteroposterior diameter changes of the rib cage and abdomen. In theory, volume displacement may be estimated from the measurement of motion of a single point on each part (Hixon, Goldman, & Mead, 1973). The distance between anterior and posterior members of a coil pair is represented instrumentally by a voltage analogue signal induced between them, which is inversely proportional to the cube of the intercoil distance (Hixon, Goldman, & Mead, 1973). When the magnetometers are initialized properly on the torso and calibrated against known volume exchanges, the voltage induced when changes in the anteroposterior diameter of the chest wall create changes in the strength of the magnetic field generated between related coils can be directly interpreted for information about the volume of air displaced by the respiratory system.

In addition to providing information about lung volume displacement, data from the magnetometers can be interpreted for information about chest wall shape and the relative contributions of the rib cage and abdomen to lung volume exchange. Thus, the muscular action operating on the chest wall to produce changes in lung volume can be inferred from a history of magnetometric data. By plotting voltage analogue signals on a relative motion chart, the experimenter can discern the unique contribution of the rib cage and abdomen to breathing events. Figure 2-2 represents a relative diameter chart, with changes in the anteroposterior diameter of the abdomen being represented on the x-axis (increasing to the right) and those of the rib cage represented on the y-axis (increasing upward). The z-axis represents volume displacement of the lungs, that being the sum of rib cage and abdominal volume represented by points on the chart (Hixon, 1987). Lung volume increases upward and rightward on the chart.

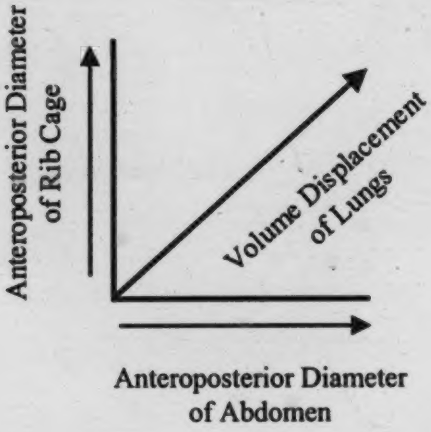
Anteroposterior Diameter of Rib Cage
Volume Displacement of Lungs
Anteroposterior Diameter of Abdomen

Figure 2- 2. Relative diameter chart.

It is possible to calibrate the z-axis for lung volume on the relative motion chart if an individual is asked to perform relaxation maneuvers against a closed airway at specified lung volumes. In addition, the functional relationship between the relative motions of the rib cage and abdomen can be revealed by means of isovolume maneuvers at those lung volumes. Figure 2-3 illustrates hypothetical isovolume pathways as four parallel diagonal lines at different lung volume levels (large black dots) in the vital capacity in relation to a relaxation characteristic (dotted line) for the chest wall in the upright body position. The isovolume pathways shown on the relative diameter chart represent maneuvers performed at four lung volume levels for which relaxation data were obtained (20, 40, 60, & 80% of vital capacity). Three speech breath groups are represented on the chart by the three lines that fall between 40 and 60% VC. Each of these lines represents speech on one expiration. The length of each line represents the volume of air exchanged during the speech produced on that breath. The slope of each line reflects the relative volume contributions of the rib cage and the abdomen to the volume exchanged. Finally, the position of each line with respect to the relaxation characteristic can be interpreted for information about the background shape of the chest wall during each expiratory breath group. Speech tracings from individuals without SCI in the upright position typically lie to the left of the relaxation curve, which indicates that the rib cage is larger and the abdomen is smaller than their relaxed configurations at the same lung volumes. Previous research indicates that individuals with SCI tend to exhibit smaller rib cage excursions and larger abdominal excursions during lung volume exchange than do individuals without SCI (Hixon & Putnam, 1983; Hoit et al., 1990). Depending on the extent of chest wall paralysis, the speech tracings of persons with SCI may lie to the right of the relaxation curve in the upright position, which indicates that the rib cage is smaller and the abdomen is larger than their relaxed configurations at the same lung volume (Hixon & Putnam, 1983; Hoit et al., 1990). Relative volume charts, such as the one in Figure 2-3, were based on the theoretical display of Figure 2-2 and used to infer chest wall shapes and chest wall part contributions to lung volume exchange during each participant's dictation in this study.

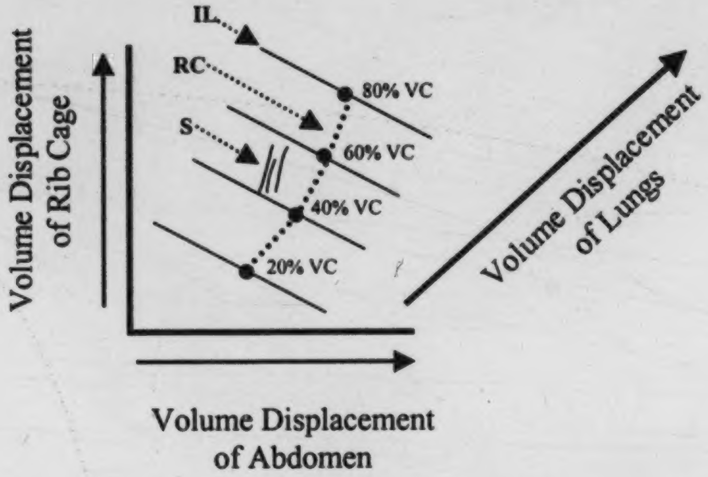
IL
RC
S
80% VC
60% VC
40% VC
20% VC
Volume Displacement of Rib Cage
Volume Displacement of Lungs
Volume Displacement of Abdomen

Figure 2- 3. Hypothetical relative volume chart for a neurologically normal speaker in the upright body position, showing isovolume lines (IL) at 20% partitions of vital capacity (VC), relaxation curve (RC), & three breath-group tracings (S).

Anteroposterior diameter changes of the rib cage and abdomen were sensed with two pairs of linearized electromagnetometers with channel frequencies of 4.2 kHz and 6.0 kHz (GMG Scientific, Inc., Burlington, MA). The magnetometer signals were sent to a power supply where they were amplified, half-wave rectified, and filtered so that the end-product was amplification of the intercoil distance change. The rib cage and abdominal magnetometer signals then were sent in parallel to a dual-channel digital storage oscilloscope (Gould, OS4200) for on-line monitoring and to separate channels of a four-channel FM instrumentation recorder (Hewlett-Packard, 3964A, San Diego, CA).

A wet respirometer (9-litre, Collins, Bionetics, Ltd., Vancouver, B.C.) was used as the volume displacement reference for the purpose of calibrating the magnetometric data. The respirometer was also used to measure each speaker's vital capacity, as well as to control volume displacements for isovolume maneuvers above the resting expiratory level. In addition, each participant performed specified inspiratory and expiratory gestures within the speech breathing range while coupled to the respirometer.

Procedure

Participants were recruited through advertisements in local media. Additionally, the director of the Canadian Paraplegic Association expressed interest in this project, and recruitment of participants with spinal cord injury was made possible through his office. Control subjects were identified for each individual with SCI after s/he had completed all experimental sessions. Control subjects were recruited primarily by word-of-mouth and were required to meet the sex, age, and height criteria of their SCI cohort before they were considered for further involvement on the study.

A meeting was arranged with each potential participant to ensure that s/he met inclusion criteria. At this meeting, the primary investigator: (1) explained what was involved in participation in the study and provided the potential participant with a copy of an information form (Appendix A), (2) judged the normalcy of each potential participant's speech and voice characteristics, (3) solicited information about pertinent health and related personal history details via a questionnaire (Appendix B), and (4) screened each potential participant's hearing and vision. If all inclusion criteria were met, interested participants were asked to sign a consent form (Appendix C). All potential participants with and without SCI who were screened met the inclusion criteria. After signing the consent form, each participant then completed three experimental sessions as follows: (1) an enrollment session, (2) a practice session, and (3) a dictation session. Each participant was reimbursed ten dollars after each session to cover transportation and parking expenses.

An enrollment session for each participant took place the same day as the initial meeting with the primary investigator. During the enrollment session (#1), each participant went through the process of training both the discrete-word and continuous-speech ASR systems to recognize his/her speech. The order in which the participants enrolled each system was randomly counterbalanced: half of the participants trained the discrete-word system first. This order was reversed for the remaining participants. Training took 30 minutes for each ASR system, and participants were provided an opportunity to rest for approximately 15 minutes between the enrollment of each system. The speech signals captured during the enrollment process were analyzed by the ASR system and stored on the computer for use in subsequent interactions with the ASR software.

During the practice session (#2), which followed the enrollment session by at least 24 hours, participants were given the opportunity to dictate using both the discrete-word and continuous-speech ASR systems. The order in which participants practiced with either system was randomly counterbalanced. Instructions were given by the primary investigator with respect to procedures for dictation and how to correct recognition mistakes. This session lasted approximately 1.5 hours.

During the dictation session (#3), which followed the practice session by at least 24 hours, data related to the respiratory and speaking variables of interest were obtained. All recording took place in a sound-treated booth where a participant was seated on a straight-backed chair or wheelchair. The transmitter coils of each surface motion magnetometer pair were attached to the front of a participant's torso at two sites using double-sided adhesive tape. The rib cage coil was placed in midline at the level of the nipples. The abdominal coil was placed in midline approximately 1 cm above the level of the umbilicus. Some individuals with spinal cord injury had distortion of their torso related to their injury, thereby making it difficult to satisfy the placement criteria on the abdomen relative to the midline. In such cases, the abdominal magnetometer was placed at the point of maximum distention of the abdomen in line with the position of the rib cage magnetometer above it, instead of in reference to the umbilicus. The respective receiver coils were placed on a participant's back in the midline at the same axial level as the transmitter coils.

Once the magnetometers were placed on each subject, the primary investigator verified that the magnetometer signals were being recorded onto channels 1 and 2 of the FM recorder. In addition, it was verified that the primary investigator's running commentary was being recorded onto channel 4 of the FM recorder via the DAT-FM microphone that was worn by the primary investigator for the preliminary breathing data collection portion of experimental session #3. Following this, the magnetometers were initialized and then each participant was asked to perform three vital capacity maneuvers. Vital capacity data provided information regarding the range of lung volume available to each speaker, as well as whether an isovolume maneuver at 1 litre above the resting expiratory level (REL) was realistic. Each participant wore a noseclip and was coupled to the respirometer via a disposable mouthpiece during these maneuvers. To accomplish them, a participant inspired freely from REL to total lung capacity (TLC) after which s/he expired fully into the respirometer. The largest of the three maneuvers performed was used to represent a participant's vital capacity. A rest period of two minutes was provided between each vital capacity maneuver. Following the vital capacity maneuvers and several minutes of quiet breathing, the magnetometers were re-initialized at REL for each participant. Thereafter, each participant performed the remaining required respiratory maneuvers and speaking tasks.

Preceding the collection of other respiratory data, the primary investigator recorded 3 minutes of resting tidal breathing from each participant while s/he sat quietly. Collection of tidal breathing data facilitated the identification of a stable baseline for REL. Throughout the remainder of the experimental recording procedure, short periods of resting tidal breathing were collected after each respiratory maneuver and speech task so that a participant's resting expiratory level could be reestablished. This was especially important when participants altered body position to relieve spasms or cramps associated with having to minimize shoulder and torso movements during magnetometric data sampling.

Following the initial collection of resting tidal breathing, each participant was asked to inspire to total lung capacity (TLC) and allow the chest wall to relax naturally down to REL via a sigh. This was done 3 times and provided information regarding chest wall shape change and volume excursion during an expiration in which the respiratory muscles were relaxed as much as a participant was able. Tracings of these sighs displayed on an oscilloscope provided an *approximation* of a relaxation curve to the REL for each speaker. Next, to verify the relaxation curve, a participant was asked to inspire to TLC, expire some air until told to hold their breath by the investigator, and then completely relax the chest wall muscles during the breath-hold. A participant was asked to do this at 3 different lung volumes levels. The investigator monitored each expiration via the magnetometric displays on the oscilloscope and instructed the participant with respect to the volume of air to be expired for each relaxation maneuver. The levels at which each participant was asked to relax against a closed airway occurred at the top of the lung capacity, near REL, and at a point roughly halfway between the two. A nose clip was worn to close the nasal airway during generation of data related to the relaxation curve, and participants were instructed to close the mouth tightly during the maneuver so that the chest wall could relax as much as possible against a closed airway. Subsequent to the collection of these relaxation data, a participant breathed quietly for 2 minutes.

Next, isovolume maneuvers were performed so that inferences about the volume displacement contribution of both the rib cage and abdomen to total lung volume exchanges could be made. For these maneuvers, a participant was asked to inspire deeply and sigh, which brought the chest wall naturally to REL. Once at REL, they held their breath and displaced volume trapped in the pulmonary system between the rib cage and abdomen by contracting and relaxing the abdominal muscles. All participants with spinal cord injury required assistance with isovolume maneuvers because of paralyzed abdominal musculature. The primary investigator assisted in displacing volume by gently pushing and then releasing the abdominal wall while the participant held his/her breath. All participants wore a noseclip and were instructed to close the mouth tightly during each isovolume maneuver. Once an isovolume maneuver was performed successfully at REL, each participant was given 2 minutes to breathe quietly in order to reestablish normal ventilation. Then, while coupled to the respirometer, a participant inspired a volume of air equivalent to approximately 20% of his/her vital capacity from REL and performed another isovolume maneuver at this level. All participants required assistance from the primary investigator who advised them when the appropriate amount of air had been inspired from the respirometer. After completion of the isovolume maneuver above REL, a participant breathed quietly for 2 minutes. Finally, for additional volume reference information, each participant exchanged 5 complete tidal breaths while coupled to the respirometer. Collection of the respiratory data thus described at the outset of session # 3 took approximately 1 - 1.5 hours per participant.

Once the preliminary respiratory data were collected, the primary investigator removed the DAT-FM microphone from her headset and snapped it into place on the headset bearing the ASR microphone. The headset bearing both microphones then was placed on a participant. The ASR microphone was positioned 2cm away from the left side of a participant's mouth as per the manufacturer's recommendations (DragonSystems, Inc., 1998). The DAT-FM microphone was positioned 2 cm away from the right side of a participant's mouth. To ensure that the ASR microphone position was adequate for speech recognition performance, a criterion rating of 'average' or better was met during the microphone setup protocol that is included with Dragon Systems' software. Simultaneous recording of the speech and speech-breathing data commenced after microphone placements were verified and signal levels optimized on the DAT and FM recorders.

Speech samples from reading aloud and dictation of a letter were elicited in each experimental condition: dictation into the tape recorder with no ASR system; dictation with a discrete-word ASR system; and dictation with a continuous-speech ASR system. The order of experimental conditions was counterbalanced. Each participant was asked to read the same stimuli and dictate a letter in each condition with a resting period of 10 minutes between each condition. Dictation with ASR was expected to be accurate, therefore, each participant used spoken commands to correct misrecognitions made by the ASR system. This was necessary in order to simulate dictation demands that users of these systems face routinely. After all three dictation conditions were completed, the magnetometers and microphone headset were removed, and the participant was thanked and excused. Collection of the speech and speech breathing data during the dictation portion of the third session took approximately 1.5 - 2 hours per participant to complete.

Hygiene Issues

The experimental protocol included steps to ensure a hygienic surrounding for each participant. When it was necessary to shave the chest and abdominal hair of a participant before placement of the magnetometers, a new disposable razor was used to do so and was discarded immediately after use. With respect to collection of respiratory data, a new disposable mouthpiece was used for each participant and was discarded immediately after use. Noseclips worn by the participants were disinfected in Wavicide-01 (Wave Energy Systems, Wayne, NJ) as per the manufacturer's specifications. The two-way breathing valve and associated hoses of the respirometer were also disinfected in Wavicide-01 as per the manufacturer's specifications.

Quiet Breathing Periods

Quiet breathing periods were required after the performance of vital capacity, relaxation, and isovolume maneuvers to allow each speaker's respiratory system to return to resting tidal inspiratory-expiratory volume exchange levels. Quiet breathing conditions also were required between speaking tasks to re-initialize the magnetometers and thereby adjust for any postural changes that may have occurred during dictation data collection. During these quiet-breathing periods, digital video clips were played for each participant to watch via the CD-ROM unit on the ASR workstation. The viewing of video clips about current events or natural history was chosen as a method to distract participants from focusing on their respiration during resting breathing periods.

Reading and Spontaneous Speech Elicitation Materials

Reading materials used to train the discrete-word ASR system during the enrollment session included pre-programmed words and short phrases found in Dragon Systems' DragonDictate software. Reading materials used to train the continuous-speech ASR system during the enrollment session consisted of one preprogrammed passage that comes as part of DragonDictate's NaturallySpeaking software. Although three preprogrammed passages are offered for enrollment in the Naturally Speaking program, one passage was chosen for this experiment to standardize the enrollment process across participants. Reading materials used during the practice session consisted of short passages that spanned approximately Grades 4-8 reading levels. Finally, reading material used during the experimental dictation session consisted of the California Passage (Appendix D), which has been specifically designed to elicit a large number of breath groups that vary in syllable number and for which normal reference data are available across a wide age range (Hoit & Hixon, 1987). For the spontaneous portion of the dictation sample, participants were asked to describe their favorite movie and to dictate the description as if they were composing a letter. The participants were given an outline to follow that provided prompts for the content of their spontaneous description (Appendix E). The outline was used to avoid possible cognitive-linguistic dysfluencies associated with extemporaneous speech without prompts (Mitchell, Hoit & Watson, 1996). After collection of each participant's data, the primary investigator transcribed the read and spontaneous speech samples in each dictation condition verbatim. Instances of filled pauses and interjections were included in the transcription.

Data Management and Analysis

Preliminary Analyses of the Data

All statistical calculations were performed using SPSS (Version 10, 1999). The distribution of the data was analyzed first to investigate its normality. The mean and median values, and the amounts of skewness and kurtosis were determined. The ratio of the skewness and kurtosis statistics to their respective standard errors can be used as a test of normality; normality is rejected if the ratio is less than -2 or greater than +2 (SPSS Applications Guide, 1999). The data sets in Table 2- 4 had values grossly outside this range.

Variable	Skewness ratio	Kurtosis ratio
Inspiratory pause time (discrete-word) (California Passage)	2.339	2.996
Number of words (no ASR) (California Passage)	5.039	8.745
Syllables per breath group (no ASR) (California Passage)	3.224	4.756
Syllables per breath group (discrete-word) (Letter)	3.923	5.807
Frequency of apnea (continuous-speech) (California Passage)	2.554	2.645
Frequency of apnea (no ASR) (Letter)	3.862	4.787

Table 2- 4. Variables for which skewness and kurtosis ratio values fell outside the range of a normal distribution.

In addition, Levene's test for homogeneity of variance was completed for all data. Review of the results of these analyses revealed that the following dependent variables had unequal variances, p<.001:

- average volume expenditure (no ASR; California Passage);
- frequency of apnea (no ASR; Letter);

Although deviations from normality and unequal variances were found, a decision to use the data without transforming them was followed for all the data in this study and is in accordance with knowledge regarding the robustness of analysis of variance statistics. With samples of equal size, the analysis of variance statistic is considered to be powerful enough to account for any reasonable departures from the assumptions of normality and homogeneity so that the validity of inferences drawn from the data are not seriously affected (Portney & Watkins, 2000).

Finally, Mauchly's test of sphericity was used to compare the covariance matrix of each dependent variable in this study with an identity matrix and thereby determine if the assumption of sphericity was met. If the assumption of sphericity was violated in any ANOVA, a Greenhouse-Geisser Epsilon correction was made. This correction adjusts the degrees of freedom used in determining the critical value of F, and makes the critical value larger. This larger value makes it harder to achieve significance, thereby compensating for the bias towards a Type I error that may be associated with unequal variances in repeated measures ANOVAs (Portney & Watkins, 2000; SPSS Applications Guide, 1999).

Reliability

Intraclass correlations (ICC) were used to determine intra-experimenter measurement reliability for all portions of this study. In addition, ICCs were used to determine interjudge reliability for the kinematic portion of the study. One of the major advantages of using this analysis is that it accounts for variance due to error components, such as judge variables. In addition, ICCs are reported to be the most generalizable measures of interjudge reliability (Portney & Watkins, 2000; Sheard, Adams & Davis, 1991; Shrout & Fleiss, 1979).

Statistical Analyses and Post-hoc Testing

The speech-breathing and speech-production portions of this study consisted of continuous variables that yielded ratio-level data. Parametric statistical procedures were used in the analyses of these data. Initially, a multiple analysis of variance (MANOVA) was considered for each portion of the study. However, because of the limited number of subjects in each group, the power in each analysis would have been significantly compromised by the number of dependent variables that would have been included within each MANOVA. Thus, several repeated-measures ANOVAs were computed for each dependent variable in the speech-breathing and speech-production data sets and their results are reported within each chapter to follow.

In order to minimize the familywise error rate of the multiple ANOVA comparisons within each portion of this investigation (i.e. the probability that at least one comparison, in a collection of comparisons, will include a Type I error), a Bonferroni correction was applied within each set of data. Bonferroni's correction is a process of adjusting alpha so that it reflects the number of comparisons within each set of data. The correction is made by dividing alpha by the number of planned comparisons within a 'family' of data (Bland, 1995; Portney & Watkins, 2000), thereby making it harder to achieve significance. Because the study was exploratory and it was important that interesting effects not be lost, the Bonferroni corrections were made for the 'families' of speech-breathing and speech-production variables separately.

Post-hoc analysis of differences between group means from each ANOVA was completed via a Bonferroni test. The Bonferroni test performs pairwise comparisons between group means, and controls overall error rate by setting the error rate for each test to the experimentwise error rate divided by the total number of tests. Thus, the observed significance level of post-hoc testing is automatically adjusted within the statistical software program for the fact that multiple comparisons are being made, so that alpha can remain at p=.05 for all post-hoc comparisons (SPSS Applications Guide, 1999).

Effect Size and Power Statistics

Effect size and power statistics were computed for the analyses of all the parametric data. Partial eta-squared will be reported as the effect size statistic in the table of results for each series of ANOVAs in the chapters that follow. The partial eta-squared statistic is an estimate of the proportion of total variability that is attributable to the factors within an experiment, and describes the amount of variance accounted for in the sample (Cohen, 1988; SPSS Applications Guide, 1999). Prospective power calculation for a 2x3x2 repeated measures design was completed using the following formula (Cohen, 1988):

$$n = \left[\frac{\left[(n_1 - 1)(u + 1)\right] + 1}{\#\,cells}\right]\#\,cells$$

With f = .8 and power = .70, the analysis revealed that a sample size of 12 would be adequate for a large effect size.

Flowchart of the Study

A flowchart is presented as a guide for the reader. The flowchart represents the experimental design applied to each set of dependent variables. In Chapter 3, the speech-breathing variables are explored between speaker groups, across dictation conditions, and between speech samples. In Chapter 4, the kinematic variables are described as they occur between speaker groups, across dictation conditions and between speech samples. In Chapter 5, the speech-production variables are considered separately for the California

Passage and for spontaneous dictation of a letter. Within each of these speech samples, the dependent variables are explored between speaker groups and across dictation conditions. A concluding chapter, Chapter 6, will summarize and integrate results from the previous chapters, and will explore issues related to threats to internal and external validity.

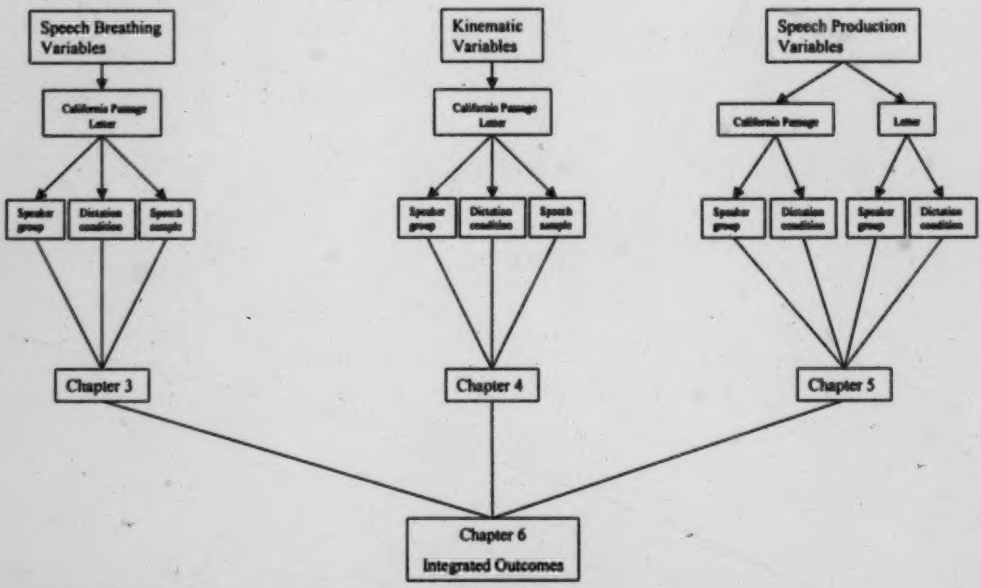
Speech Breathing Variables
California Passage Letter
Speaker group
Dictation condition
Speech sample
Chapter 3
Kinematic Variables
California Passage Letter
Speaker group
Dictation condition
Speech sample
Chapter 4
Speech Production Variables
California Passage
Letter
Speaker group
Dictation condition
Speaker group
Dictation condition
Chapter 5
Chapter 6
Integrated Outcomes

Figure 2- 4. Flowchart outlining the experimental design for each subset of analyses.

CHAPTER 3 - Speech Breathing Results

Purpose

The intent of this portion of the study was to describe the characteristics of breathing patterns associated with spontaneous speech and reading aloud in three dictation conditions (with discrete-word ASR, with continuous-speech ASR, and with no ASR), and to determine if breathing characteristics differed between two groups of users (those with and without spinal cord injury) and two speech samples (spontaneous and scripted speech).

Variables

The dependent variables for this portion of the study were as follows:

a. Mean volume (%VC) per pre-speech inspiration

b. Mean volume (%VC) per expiratory segment for speech

c. Average absolute volume expenditure per expiratory segment for speech (ml)

d. Duty cycle of the respiratory system within the expiratory phase for speech (% speech)

e. Average duration of inspiratory pause time within a given passage (ms)

Hypotheses & Hypothetical Outcomes

Hypotheses and related hypothetical outcomes that were proposed for the breathing variables of interest are reproduced in Table 3-1.

Dependent Variable	Speaker Group	Dictation Condition	Speech Sample
Mean **volume** (%VC) per pre-speech inspiration across a given passage…	will be higher for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be lowest in the discrete-ASR condition and highest in the no-ASR condition for all speakers and speech samples	will be higher for reading than for spontaneous speech for all speakers and dictation conditions
Mean **volume** (%VC) per expiratory segment for speech across a given passage…	will be higher for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be lowest in the discrete-ASR condition and highest in the no-ASR condition for all speakers and speech samples	will be higher for reading than for spontaneous speech for all speakers and dictation conditions
Average absolute **volume expenditure** per expiratory segment for speech (ml) across a given passage…	will be smaller for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be smallest when using discrete-word ASR and largest in the no-ASR dictation condition for all speakers and speech samples	will be the same for both reading and spontaneous speech for all speakers and dictation conditions
Duty cycle of the respiratory system within the expiratory phase for speech (% speech)…	will be higher for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be lowest when using discrete-word ASR and highest in the no-ASR dictation condition for all speakers and speech samples	will be higher for reading than for spontaneous speech for all speakers and dictation conditions
Average duration of **inspiratory pause time** (ms) within a given passage…	will be longer for individuals with SCI than for those without SCI across dictation conditions and between speech samples	will be shortest when using discrete-word ASR and longest in the no-ASR dictation condition for all speakers and speech samples	will be shorter for reading than for spontaneous speech for all speakers and dictation conditions

Table 3- 1. Hypothetical outcomes for speech-breathing variables between speaker groups, across dictation conditions, and between speech samples.

Method

Data Processing and Analysis

Both a dual-channel digital storage oscilloscope (Gould, OS4200) and an acoustic analysis software package (CSpeech, Version 4; Milenkovic, 1992; Madison, Wisconsin) were used in the processing and analysis of the speech-breathing data in this portion of the study. For three of the dependent variables (mean volume per pre-speech inspiration, mean volume per expiratory segment for speech, & average absolute volume expenditure per expiratory segment for speech), the multichannel data associated with the dictation conditions on FM tape (i.e., the RC magnetometer and ABD magnetometer signals) were sent to the oscilloscope for analysis by the primary investigator (See Figure 3-1). For two of the dependent variables (average duration of inspiratory pause time & duty cycle of the respiratory system within the expiratory phase for speech), the multichannel data associated with the dictation conditions on FM tape were digitized and analyzed using CSpeech. The rib cage, abdomen, and speech signals from the FM record were low pass filtered at 5kHz via matched 8-pole Butterworth lowpass filters (Frequency Devices 901; Haverhill, MA), and digitized at 11kHz and a 12-bit quantization rate (Data Translation 2821; Vancouver, B.C.). For this analysis, the speech signal was used only as a boundary marker for respiratory events; thus, it was unnecessary to digitize the speech waveform at a higher sampling rate. The information extracted from the CSpeech software program can be likened to that obtained via an oscilloscope in the TY mode with the added benefits of measurement functions tailored to speech waveforms and playback capabilities for them. The three waveform traces were displayed simultaneously in the same window as shown in Figure 3-2.

For all speech breathing data, 30 respiratory cycles that supported speech were measured for each dependent variable. Within the California Passage (Appendix D), the 30 respiratory cycles for speech consisted of the 3rd through the 13th (disregarding the first 3 respiratory cycles for speech to avoid any onset effects), followed by the 10 respiratory cycles for speech that began at the phrase, "In the winter…", and ending with the 10 respiratory cycles for speech that preceded the last 3 (which were disregarded to avoid any offset effects). During dictation of the California Passage with no ASR, there were often less than 30 respiratory cycles that supported speech. In these cases, all respiratory cycles that supported speech were measured except the first and last three. During spontaneous dictation of a letter, the 30 respiratory cycles for speech that followed the first 3 (which were disregarded) were measured.

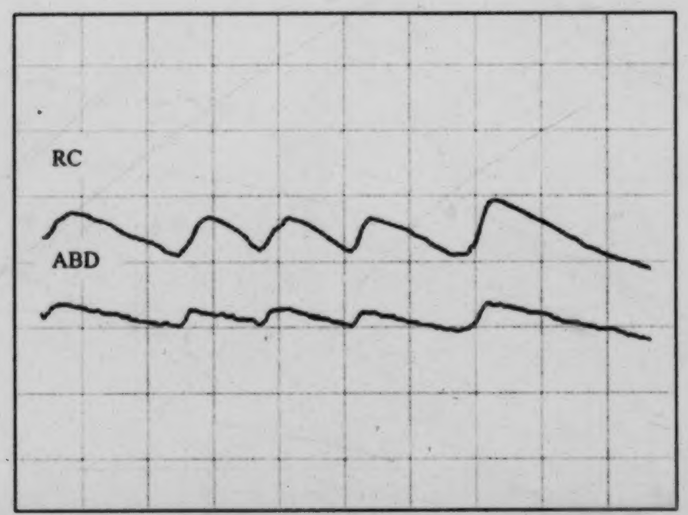

Figure 3- 1. A schematic of the rib cage (RC) and abdomen (ABD) magnetometric waveforms for 5 utterance cycles as they appeared on the oscilloscope screen.

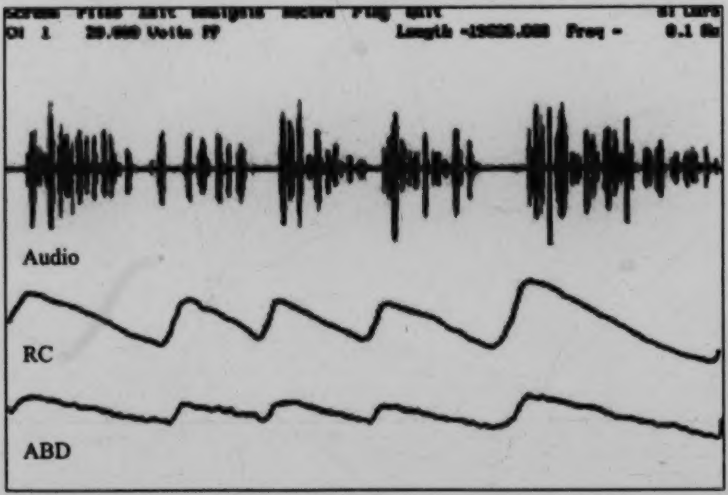

Figure 3- 2. Audio, rib cage (RC), and abdomen (ABD) waveforms for 5 respiratory cycles that supported speech as displayed in CSpeech for a participant dictating with no ASR. The x-axis in the figure represents time (sec), and the y-axis represents amplitude (volts). Inspiratory portions of the breathing cycles are up-going moments on the RC and ABD traces; expiratory portions are down-going.

Prior to measurement of inspiratory and expiratory volume data, the primary investigator calculated the contribution of the rib cage and abdomen to lung volume exchange in reference to the tidal breaths that each participant performed on the respirometer during preliminary collection of respiratory data. A referent in volts was calculated for units of volume exchanged. The sum of the rib cage and abdomen's voltages for each corresponding tidal breath was compared to actual volume displaced on the respirometer chart paper to verify the relationship between chest wall part contributions to lung volume exchange and their voltage analogues. A BTPS (body temperature pressure saturation) correction factor was applied to the volumes displaced on the respirometer chart paper. Through a simple mathematical equation, total volume displaced per volt was calculated for the inspiratory and expiratory portions of each tidal breath, and mean volume displaced per volt was calculated across the 5 tidal breaths. The mean volume displaced was entered into a spreadsheet and was used in conversion of voltages to volumes for the dependent variables, mean volume per pre-speech inspiration, mean volume per expiratory segment for speech, and average absolute volume per expiratory segment for speech.

Mean Volume (%VC) per Pre-Speech Inspiration

Inspiratory volume was determined by locating and measuring the inspiratory gestures for 30 breaths that supported speech in the data displayed on the oscilloscope screen (see Figure 3-3). In the T-Y mode, the inspiratory portion was that part of the trace in which the slope changed upward (Figure 3-3, line A-B). The extents of the RC and ABD inspiratory traces were measured in volts (Figure 3-3, point B = end-inspiratory level for RC), summed and then converted to units of volume inspired. These volumes were transformed to a percentage of each participant's vital capacity in order to allow comparison across individuals. The average inspiratory volume (%VC) over the 30 inspirations was calculated and used in the statistical analysis.

Mean Volume (%VC) per Expiratory Segment for Speech

Expiratory volume was determined by locating and measuring the expiratory portions of 30 respiratory cycles that supported speech in the data displayed on the oscilloscope screen (see Figure 3-3). In the T-Y mode, the expiratory portion was that part of the trace in which the slope changed downward to indicate expiration (Figure 3-3, line B-C). The extents of the RC and ABD expiratory traces were measured in volts (Figure 3-3, point C = end-expiratory level for RC), summed and then converted to units of volume expired. These volumes were transformed to a percentage of each participant's vital capacity in order to allow comparison across individuals. The average of the expiratory volumes (%VC) over the 30 expiratory segments for speech was calculated and used in the statistical analysis.

Average Absolute Volume Expenditure per Expiratory Segment for Speech

Average absolute volume expenditure per expiratory segment for speech (Figure 3-3, line B-C) was recorded for the same data that were transformed to express mean expiratory volume (%VC) per expiratory segment for speech. The mean absolute volume exchanged across the 30 expiratory segments for speech was thus included in the statistical analyses, as well.

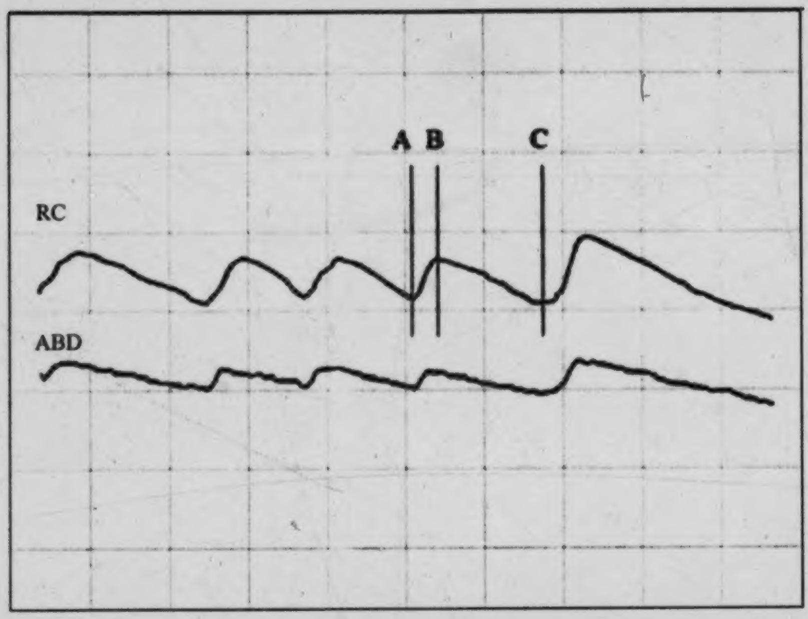

Figure 3- 3. A schematic of the rib cage (RC) and abdomen (ABD) magnetometric waveforms for 5 utterance cycles as they appeared on the oscilloscope screen. On the rib cage trace, Point A = initiation of inspiration; Point B = end-inspiratory level; Point C = end-expiratory level; line B-C = expiratory volume expenditure related to the RC. In this figure, the x-axis represents time (sec), and the y-axis represents amplitude (volts).

Duty Cycle of the Respiratory System within the Expiratory Phase for Speech

The duty cycle of the respiratory system within the expiratory phase for speech was determined in CSpeech by locating and measuring the expiratory segments associated with 30 breath groups. Within a window bearing speech and respiratory waveforms, the beginning of an expiration on the magnetometric waveforms was marked with one cursor (Figure 3-4; line B) and the end of expiration was marked with another cursor (Figure 3-4; line D). The time between cursors (B-D) was recorded. Then, for the same expiratory segment, one cursor was placed at the beginning of the acoustic waveform (Figure 3-4; line B), and another cursor was placed at the end of the acoustic waveform (Figure 3-4; line C). The time between cursors (B-C) was recorded. From these two measurements (i.e., total expiratory time versus time spent speaking), the percentage of the expiration that contained speech was calculated, and the mean across the 30 measurements was computed.

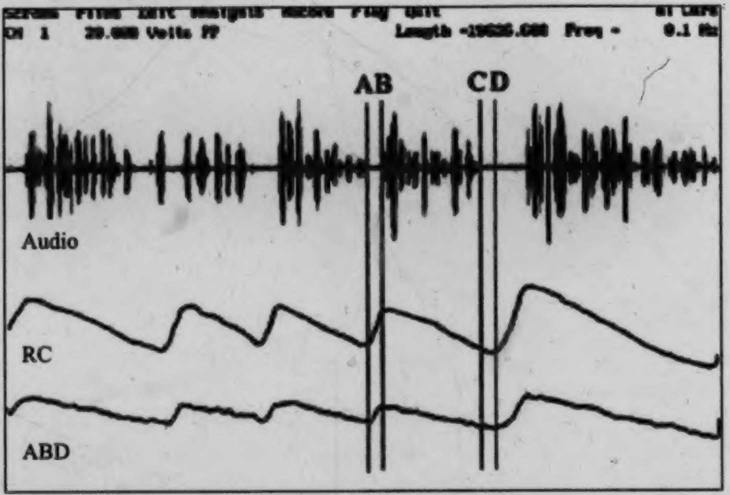

Figure 3- 4. Audio, rib cage and abdomen waveforms for 5 utterance cycles as viewed in CSpeech. Line A-B = inspiratory pause time; Line B-C = speech produced within the expiratory phase; Line B-D = the total expiratory phase. In this figure, the x-axis represents time (sec), and the y-axis represents amplitude (volts).

Average Duration of Inspiratory Pause Time

Average duration of inspiratory pause time per pre-speech inspiration was determined in CSpeech by locating and measuring the inspiratory portions (Figure 3-4; line A-B) of 30 respiratory cycles that supported speech. The time between the cursors at points A and B was recorded, and the mean across the 30 measurements was computed.

Statistical Analyses

The speech-breathing variables were analyzed via a set of 2x2x3 (speaker group x speech sample x dictation condition) repeated-measures ANOVAs with fixed effects on the first factor and repeated-measures on the last two factors. Each dependent variable was analyzed via a separate ANOVA, each having 2 levels on the first factor (SCI & non-SCI), 2 levels on the second factor (California Passage & Letter), and 3 levels on the third factor (no ASR, discrete-word ASR, & continuous-speech ASR). Reliability for the measurement of the dependent variables was analyzed via intra-class correlations (ICC) for approximately 25% of the data collected.

Results

Table 3-2 contains mean scores and standard deviations obtained for all the dependent variables for speech breathing between groups, across the three dictation conditions, and between speech samples.

		dictation condition	speech sample	Mean	Std. Deviation
mean volume (%VC) per pre-speech inspiration	SCI	discrete-word	California	12.49	4.61
			Letter	14.0[illegible]	7.46
		continuous-speech	California	17.94	6.96
			Letter	16.45	7.38
		no ASR	California	17.56	5.69
			Letter	18.02	5.31
	non-SCI	discrete-word	California	10.41	5.94
			Letter	11.14	7.25
		continuous-speech	California	11.86	7.20
			Letter	12.04	7.86
		no ASR	California	15.58	9.96
			Letter	13.87	9.59
mean volume (%VC) per expiratory segment for speech	SCI	discrete-word	California	12.55	4.75
			Letter	14.75	8.26
		continuous-speech	California	18.19	7.14
			Letter	16.54	7.34
		no ASR	California	17.55	5.74
			Letter	18.29	5.27
	non-SCI	discrete-word	California	10.40	6.06
			Letter	10.88	6.85
		continuous-speech	California	12.85	7.58
			Letter	12.36	8.30
		no ASR	California	15.71	9.91
			Letter	13.70	9.55
average volume expenditure (ml) per expiratory segment for speech	SCI	discrete-word	California	269.7	119.70
			Letter	302.6	132.60
		continuous-speech	California	381.9	137.40
			Letter	345.2	133.30
		no ASR	California	367.4	92.90
			Letter	393.2	138.50
	non-SCI	discrete-word	California	397.3	177.10
			Letter	413.4	208.30
		continuous-speech	California	489.6	234.00
			Letter	488.0	249.20
		no ASR	California	598.7	332.00
			Letter	517.5	292.80
duty cycle of the respiratory system (%speech)	SCI	discrete-word	California	56.39	6.19
			Letter	53.13	9.77
		continuous-speech	California	65.64	13.75
			Letter	58.01	5.52
		no ASR	California	93.94	3.77
			Letter	87.05	8.14
	non-SCI	discrete-word	California	48.20	11.88
			Letter	44.50	7.15
		continuous-speech	California	58.30	10.22
			Letter	49.27	7.57
		no ASR	California	96.51	2.23
			Letter	86.57	6.54
inspiratory pause time (msec)	SCI	discrete-word	California	773.0	225.55
			Letter	796.5	150.45
		continuous-speech	California	826.4	171.87
			Letter	862.9	144.56
		no ASR	California	504.8	69.04
			Letter	642.2	121.09
	non-SCI	discrete-word	California	791.6	73.21
			Letter	812.1	172.12
		continuous-speech	California	841.7	87.28
			Letter	839.4	124.25
		no ASR	California	549.4	87.64
			Letter	581.5	82.69

Table 3- 2. Mean scores and standard deviations obtained for all the speech-breathing variables between groups, across the three dictation conditions, and between speech samples.

Within-subjects results from each repeated-measures ANOVA for the speech-breathing variables are summarized in Table 3-3. Between-groups results are summarized in Table 3-4. Alpha levels were adjusted from p = .05 to p = .01 according to Bonferroni's correction for 5 planned comparisons (.05/5) (Portney & Watkins, 2000). Significant main effects were found within dictation condition and speech sample for several of the speech-breathing variables. No interaction effects were found. In addition, no significant differences were found between groups for any of the respiratory variables. Because no group differences were found, the results that are illustrated in Figures 3-5, 3-8, 3-10, 3-12 and 3-14 represent main effects for dictation condition averaged across all speakers. For the reader who is interested in group means, please refer to Table 3-2.

Univariate Tests

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared	Observed Power
DICTATION CONDITION	inspiratory volume	Sphericity Assumed	217.271	2	108.636	16.396	.000*	.621	.999
	expiratory volume	Sphericity Assumed	217.628	2	108.814	18.293	.000*	.647	.999
	average volume	Greenhouse-Geisser	186040.462	1.215	153067.745	17.417	.001*	.635	.993
	duty cycle	Sphericity Assumed	22342.285	2	11171.142	114.783	.000*	.920	1.000
	inspiratory pause time	Sphericity Assumed	1016858.352	2	508429.176	37.176	.000*	.788	1.000
DICTATION CONDITION * GROUP	inspiratory volume	Sphericity Assumed	25.118	2	12.559	1.895	.176	.159	.347
	expiratory volume	Sphericity Assumed	10.947	2	5.474	.920	.416	.084	.186
	average volume	Sphericity Assumed	1.472E-02	2	7.362E-03	1.379	.275	.121	.262
	duty cycle	Sphericity Assumed	343.944	2	171.972	1.767	.196	.150	.326
	inspiratory pause time	Sphericity Assumed	2187.377	2	1093.688	.080	.923	.008	.061
SPEECH SAMPLE	inspiratory volume	Sphericity Assumed	2.726E-02	1	2.726E-02	.005	.944	.001	.050
	expiratory volume	Sphericity Assumed	.272	1	.272	.049	.829	.005	.055
	average volume	Sphericity Assumed	2.007E-03	1	2.007E-03	.374	.555	.036	.086
	duty cycle	Sphericity Assumed	817.457	1	817.457	14.879	.003*	.598	.934
	inspiratory pause time	Sphericity Assumed	30703.232	1	30703.232	1.659	.227	.142	.215
SPEECH SAMPLE * GROUP	inspiratory volume	Sphericity Assumed	.954	1	.954	.180	.680	.018	.067
	expiratory volume	Sphericity Assumed	5.515	1	5.515	.991	.343	.090	.147
	average volume	Sphericity Assumed	5.914E-03	1	5.914E-03	1.054	.329	.095	.154
	duty cycle	Sphericity Assumed	12.004	1	12.004	.218	.650	.021	.071
	inspiratory pause time	Sphericity Assumed	10814.472	1	10814.472	.584	.462	.055	.107
DICTATION CONDITION * SAMPLE	inspiratory volume	Sphericity Assumed	13.135	2	6.567	1.479	.252	.129	.278
	expiratory volume	Sphericity Assumed	19.708	2	9.854	1.983	.164	.165	.361
	average volume	Sphericity Assumed	1.122E-02	2	5.609E-03	1.246	.309	.111	.240
	duty cycle	Sphericity Assumed	95.880	2	47.940	1.191	.324	.106	.231
	inspiratory pause time	Sphericity Assumed	17061.101	2	8530.551	.637	.539	.060	.142
DICTATION CONDITION * SPEECH SAMPLE * GROUP	inspiratory volume	Sphericity Assumed	11.415	2	5.708	1.285	.298	.114	.246
	expiratory volume	Sphericity Assumed	12.257	2	6.128	1.234	.312	.110	.238
	average volume	Sphericity Assumed	1.206E-02	2	6.027E-03	1.339	.285	.118	.255
	duty cycle	Sphericity Assumed	5.276	2	2.638	.066	.937	.007	.060
	inspiratory pause time	Sphericity Assumed	8090.074	2	4045.037	.302	.743	.029	.091

Table 3- 3. Summary of ANOVA results for within-subjects results for each speech-breathing variable. *Indicates a statistically significant result.

Transformed Variable: Average

Source	Measure	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared	Observed Power
GROUP	inspiratory volume	234.364	1	234.364	.812	.389	.075	.129
	expiratory volume	241.452	1	241.452	.803	.391	.074	.128
	average volume	339878.946	1	339878.946	1.568	.239	.136	.206
	duty cycle	474.519	1	474.519	5.418	.042	.351	.556
	inspiratory pause time	51.764	1	51.764	.001	.970	.000	.050

Table 3- 4. Summary of ANOVA results for between-groups results for each speech-breathing variable.

Mean Volume (%VC) per Pre-Speech Inspiration

The analysis revealed significant differences in mean volume per pre-speech inspiration among dictation conditions. Results for mean inspiratory volumes as a percentage of the vital capacity are illustrated on the left side of Figure 3-5. Inspection of the figure reveals that subjects inspired the highest percentage of their vital capacity when dictating with no ASR, and the lowest percentage of their vital capacity when dictating with discrete-word ASR (16.25% & 12.01%, respectively). In the continuous-speech ASR condition, subjects inspired a percentage of the VC that fell between the other two dictation conditions (14.57%). Bonferroni post-hoc analyses revealed that mean volume per pre-speech inspiration differed significantly between dictation with discrete-word ASR and dictation with both continuous-speech ASR and no ASR, p<.01. There were no significant differences between dictation with continuous-speech ASR and no ASR, p=.127. Partial eta-squared values revealed that 62% of the variance in mean inspiratory volume was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 3-6.

Mean Volume (%VC) per Expiratory Segment for Speech

Results for findings related to volumes expended per expiratory segment for speech as a percentage of the vital capacity are illustrated on the right side of Figure 3-5. Subjects expired the highest percentage of their vital capacity when dictating with no ASR, and the lowest percentage when dictating with discrete-word ASR (16.31% and 12.14%, respectively). In the continuous-speech ASR condition, subjects expired a percentage of the vital capacity that fell between the other two dictation conditions (14.98%). Bonferroni post-hoc analyses revealed that percentage of vital capacity expired for dictation with discrete-word ASR differed significantly from the percentage exchanged during dictation with both continuous-speech ASR and no ASR, p<.01. There were no significant differences in expiratory volumes between dictation with continuous-speech ASR and no ASR, p=.138. Partial eta-squared values revealed that 65% of the variance in mean expiratory volume per expiratory segment for speech was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 3-7.

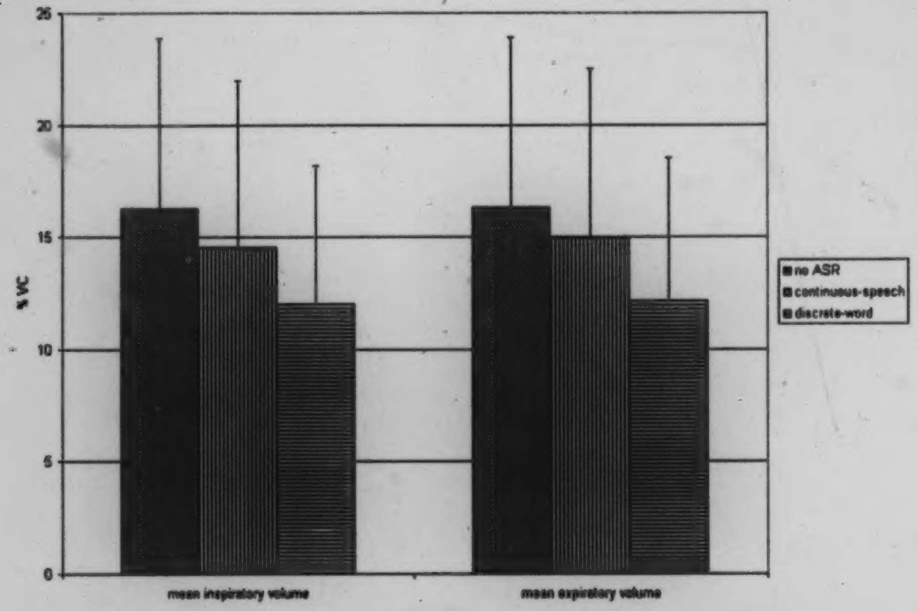

Figure 3- 5. Main effects for both mean inspiratory and mean expiratory volumes expressed as percent vital capacity across dictation conditions. For these speech-breathing variables, differences between the discrete-word ASR condition and both the continuous-speech and no-ASR conditions were significant. There was no significant difference between the continuous-speech ASR and the no-ASR conditions.

Figure 3- 6. Partial eta-squared values for mean volume per pre-speech inspiration. The proportion of total variability in the sample attributable to dictation condition was 62%.

Figure 3- 7. Partial eta-squared values for mean volume per expiratory segment for speech. The proportion of total variability in the sample attributable to dictation condition was 65%.

Average Absolute Volume Expenditure per Expiratory Segment for Speech

Results for the main effects found for average absolute volume expenditure (ml) per expiratory segment for speech across dictation conditions are illustrated in Figure 3-8. Examination of these data reveal that subjects expended the greatest amount of air per expiratory segment for speech in the no ASR condition (469 ml), followed by the continuous-speech (421 ml) and discrete-word (346 ml) conditions. Bonferroni post hoc analyses revealed that average absolute volume expenditure per expiratory segment for speech during dictation with discrete-word ASR differed significantly from average absolute volume expenditure during dictation with both continuous-speech ASR and no ASR, $p<.01$. There were no significant differences in average absolute volume expenditure between dictation with continuous-speech ASR and no ASR, $p=.068$. Partial eta-squared values revealed that 63% of the variance in average absolute volume expenditure per expiratory segment for speech was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 3-9.

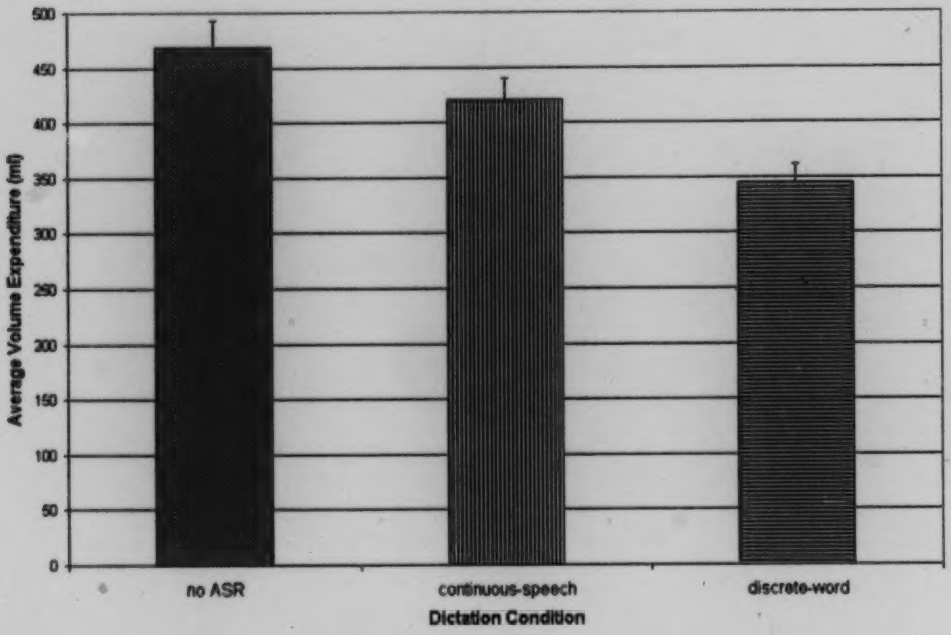

Figure 3- 8. Main effects for average absolute volume expenditure per expiratory segment for speech across dictation conditions. Differences between the discrete-word ASR condition and both the continuous-speech ASR and no-ASR conditions were significant. There was no significant difference between the continuous-speech ASR and the no-ASR conditions.

Figure 3- 9. Partial eta-squared values for average absolute volume expenditure per expiratory segment for speech. The proportion of total variability in the sample attributable to dictation condition was 63%.

Duty Cycle of the Respiratory System within the Expiratory Phase for Speech

Differences between the means for the duty cycle of the respiratory system within the expiratory phase for speech (% speech) in each dictation condition are shown in Figure 3-10. When dictating with no ASR, subjects used a greater proportion of their expiratory breaths for speech (91.02%) than they did when dictating with either continuous-speech ASR (57.81%) or discrete-word ASR (50.56%). Bonferroni post-hoc analyses revealed that these differences were significant between the no-ASR dictation condition and both the discrete-word and continuous-speech ASR conditions, p<.0001. There was no significant difference between the duty cycle characteristics of discrete-word ASR and continuous-speech ASR, p=.152. Partial eta-squared values revealed that 92% of the variance in duty cycle of the respiratory system within the expiratory phase for speech was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 3-11.

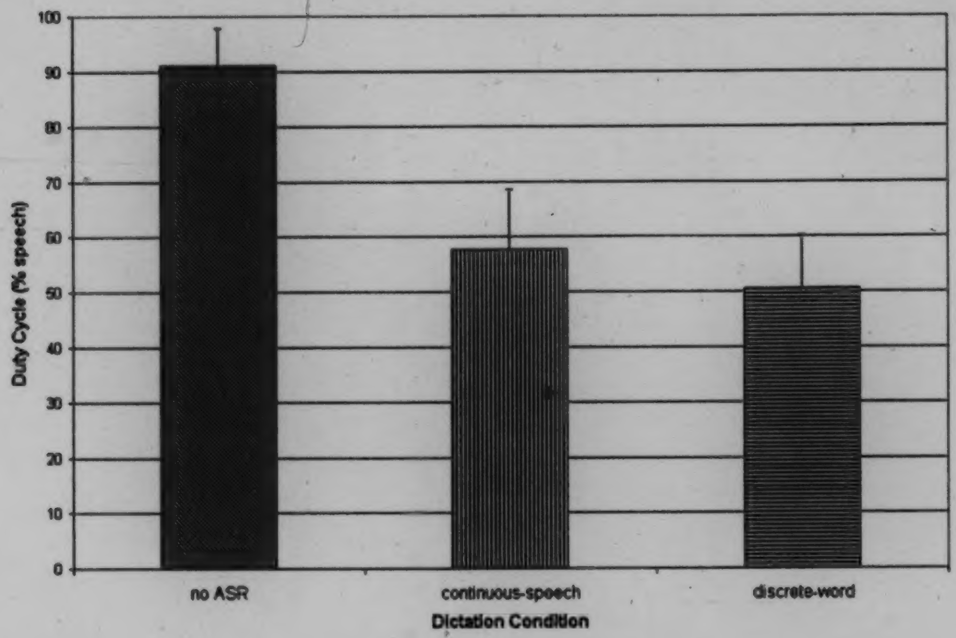

Figure 3- 10. Main effects for duty cycle of the respiratory system within the expiratory phase for speech across dictation conditions. Differences between the no-ASR condition and both the continuous-speech and discrete-word ASR conditions were significant. There was no significant difference between the continuous-speech and discrete-word ASR conditions.

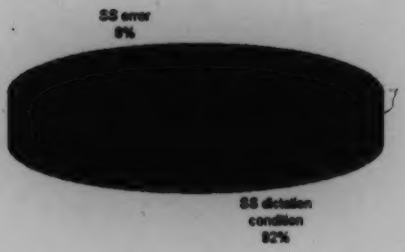

Figure 3- 11. Partial eta-squared values for duty cycle of the respiratory system within the expiratory phase for speech. The proportion of total variability in the sample attributable to dictation condition was 92%.

In addition, significant main effects were found between speech samples for duty cycle of the respiratory system within the expiratory phase for speech. These results are illustrated in Figure 3-12. When subjects dictated the California Passage, they used a greater proportion of each expiratory breath for speech (69.83%) than they did when dictating a letter spontaneously (63.09%), p<.01. Partial eta-squared values revealed that 60% of the variance in duty cycle of the respiratory system within the expiratory phase for speech was attributable to speech sample. The remaining proportion of the variance was due to error. This is illustrated in Figure 3-13.

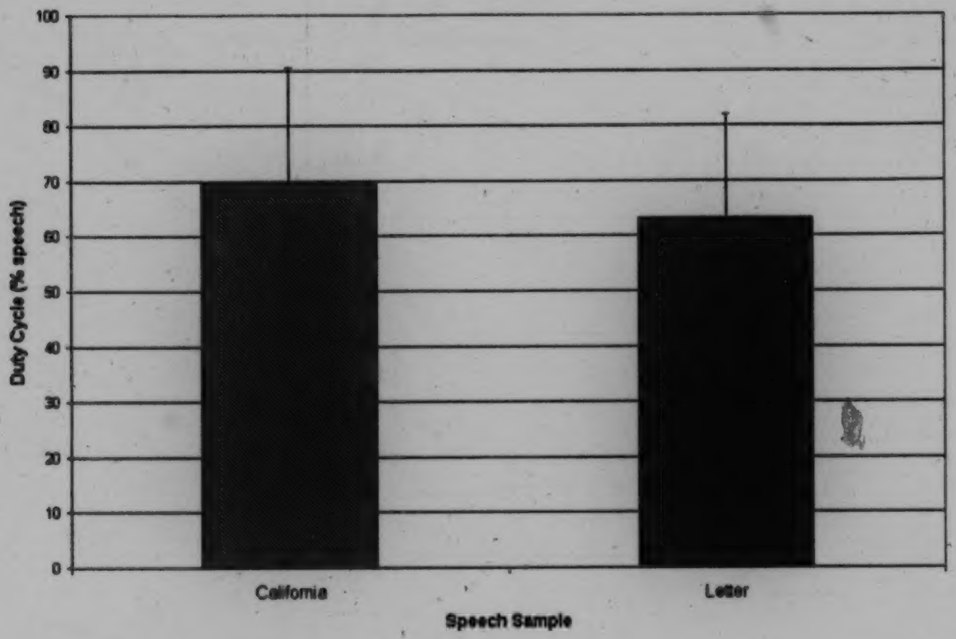

Figure 3- 12. Main effects found for duty cycle of the respiratory system within the expiratory phase for speech between speech samples.

Figure 3- 13. Partial eta-squared values for duty cycle of the respiratory system within the expiratory phase for speech. The proportion of total variability in the sample attributable to speech sample was 60%.

Average Duration of Inspiratory Pause Time

Differences existed between the means for inspiratory pause time (msec) in each dictation condition and are shown in Figure 3-14. When dictating with no ASR, subjects spent less time in the inspiratory phase for speech (570 msec) than they did when dictating with either continuous-speech ASR (843 msec) or discrete-word ASR (793 msec). Bonferroni post-hoc analyses revealed that differences in inspiratory pause time were significant between the no-ASR dictation condition and both the discrete-word and continuous-speech ASR conditions, p=.001. Differences in inspiratory pause times between discrete-word ASR and continuous-speech ASR were not significant, p=.464. Partial eta-squared values revealed that 79% of the variance in inspiratory pause time was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 3-15.

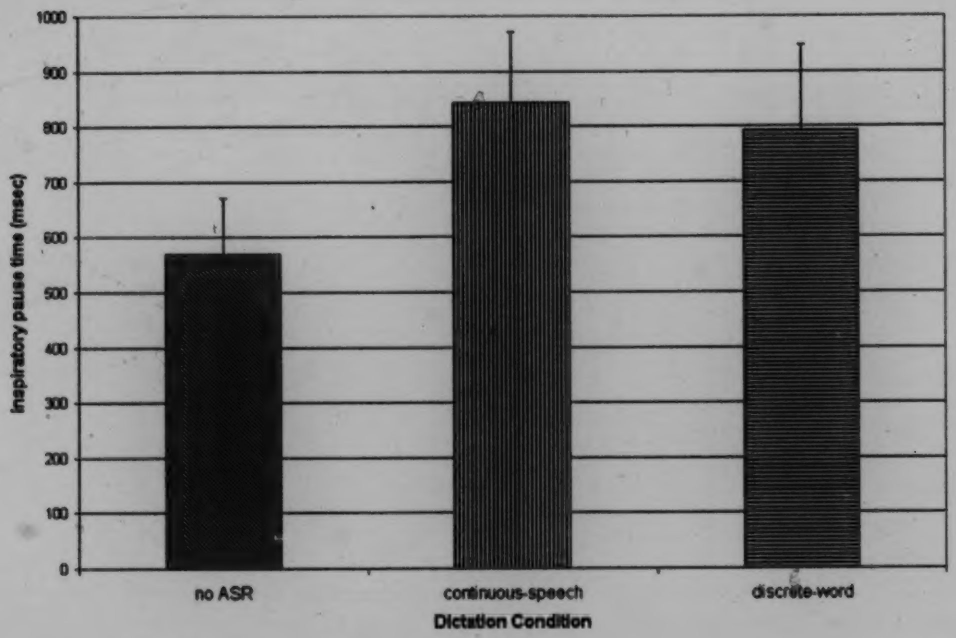

Figure 3- 14. Main effects for inspiratory pause time across dictation conditions. Differences between the no-ASR condition and both the continuous-speech and discrete-word ASR conditions were significant. There was no significant difference between the continuous-speech and the discrete-word ASR conditions.

Figure 3- 15. Partial eta-squared values for inspiratory pause time. The proportion of total variability in the sample attributable to dictation condition was 79%.

Measurement Reliability

Measurement reliability was assessed via intra-class correlations (ICC) for 25% of the breathing data collected in this portion of the investigation. The ICC coefficient for reliability was 0.8589. A confidence interval of 95% was established with a lower bound of .8320 and an upper bound of .8816, which encompasses the ICC coefficient for reliability obtained for these repeated measurements. Mean measurement error for volume-related data was 16.36 ml and for %VC measurements was 0.77%. Mean measurement error for duty cycle was 7.91% and for inspiratory pause time was 13.86 msec.

Discussion

The intent of this portion of the investigation was to determine if speech-breathing behaviors differed across three dictation conditions in which individuals used discrete-word ASR, continuous-speech ASR and no ASR to dictate a scripted paragraph and a spontaneous letter. Speech produced in the no-ASR condition will be referred to as the control condition and the individuals not affected by spinal cord injury will be referred to as the control group. The results of this portion of the experiment will be discussed in the text that follows with respect to other studies of both individuals with SCI and their non-disabled peers reported in the literature, and to the hypotheses posed for each speech-breathing variable across dictation conditions, between speaker groups, and between speech samples.

Mean Volume (%VC) per Pre-Speech Inspiration

Mean volume per pre-speech inspiration was the amount of air that was inspired prior to an expiration that supported speech, expressed as a percent of the speaker's vital capacity. The %VC form of the variable was used so that comparisons could be made across individuals. The pre-speech inspiration volumes exhibited by the control group in the present study for dictation in the control condition ranged from 13.87 to 15.58 %VC across both speech samples. This range is consistent with that reported in other studies of non-disabled individuals (Winkworth & Davis, 1997; Winkworth, Davis, Adams & Ellis, 1995; Winkworth, Davis, Ellis & Adams, 1994). Winkworth and Davis (1997) report a pre-speech volume inspiration range of 12.74 to 19.25 %VC across reading aloud and monologue tasks. Winkworth and colleagues (1995) report pre-speech volume inspiration of 12.68 %VC during spontaneous speech, and 11.31 %VC during reading (Winkworth et al., 1994).

Across the experimental variables in the current study, significant results for %VC inspired for speech were found for dictation condition only. Results for speaker group and speech sample were nonsignificant. Percent VC inspired for speech was predicted to be lower in the discrete-word ASR condition than in the control condition. This prediction was based on expectations about the effects of ASR software on speech and results of pilot-study work. The %VC inspired was hypothesized to be greatest when dictating in the control condition because of well-documented demands for respiratory support during natural connected speech. These demands include longer breath groups and the influence of linguistic factors, such as sentence boundaries, that have a bearing on the location of inspirations (Winkworth, Davis, Adams & Ellis, 1995; Winkworth, Davis, Ellis & Adams, 1994). Because use of discrete-word ASR requires pauses between each word, it was expected that shorter breath groups and smaller %VC pre-speech inspirations would characterize its use. In contrast, pre-speech inspiratory volumes were expected to represent more of the vital capacity when subjects were dictating with continuous-speech ASR than when they dictated with discrete-word ASR. Continuous-speech ASR is touted to allow more natural speech production, reflected in longer breath groups than those observed during dictation with discrete-word dictation, and juncture that more closely approximates the interword boundary characteristics of natural connected speech.

The hypothesis developed for differences across dictation conditions for this dependent variable was supported by the findings from this study, however the outcomes of the post-hoc analyses were not all significant. Differences were significant only between dictation with discrete-word ASR and the other two dictation conditions. Although volume per pre-speech inspiration was less in the continuous-speech condition than in the control condition, the difference was not significant. This suggests that discrete-word ASR constrains pre-speech inspiratory behavior unnaturally, but that continuous-speech ASR does not constrain inspirations to a degree that differs from normal speech. This finding is supported by results reported by Winkworth and colleagues (1995) with respect to depth of inspiration and the influence of linguistic factors. These authors found that inspirations that occurred at what they termed structural boundaries (i.e., before a clause that contained a finite verb) were larger than inspirations that occurred at phrase boundaries, between lists of items, or before single words or filler words. Thus, because natural speech and dictation with continuous-speech ASR is more likely to be characterized by breath groups that consist of at least one clause or several clauses, those speaking conditions are likely to be associated with deeper inspirations. Dictation with discrete-word ASR, on the other hand, is more likely to be characterized by single-word input, and therefore the respiratory system adjusts for those expectations.

It appears then that the speech-breathing mechanism is regulated during dictation with ASR in a pattern that reflects the type of speech input required for successful dictation.

A priori, it also had been hypothesized that the % VC inspired for speech would be larger in individuals with SCI than in non-disabled individuals. This hypothesis was based on reports documenting the limitations that SCI imposes on respiratory capacities (Hoit, Banzett, Brown & Loring, 1990), and on previous work that suggested differences in breathing for speech between individuals with and without SCI (Hixon & Putnam, 1983). Hoit and colleagues (1990) have described reductions in vital capacity and inspiratory capacity in individuals with SCI, which can result in pre-speech inspiration values for these individuals that represent a larger percentage of their vital capacity than that of their non-disabled cohorts.

The results for %VC inspired between groups within the present study confirm the direction of the proposed hypothesis. Individuals with SCI inspired a greater proportion of their vital capacity for speech than did individuals without SCI (16 %VC versus 12% VC, respectively). However, the ANOVA results revealed that differences in %VC inspired for speech between the two groups were not significant. A lack of significant differences between groups differs from results reported by Hixon and Putnam (1983). They described a difference in percent vital capacity volume utilized for pre-speech inspiration excursions between non-disabled speakers and an individual with SCI. In this case report, Hixon and Putnam found that their injured subject's volume excursions covered approximately 5 to 10 percent of his predicted vital capacity, whereas non-disabled subjects typically exhibited pre-speech inspirations that covered 10 to 20 percent of their *predicted* vital capacities. That no significant group differences were found in the present study like those described by Hixon and Putnam may be due to the fact that the percent vital capacity measures in this investigation were computed from the subjects' *actual* vital capacity measurements and not from their predicted values. The predicted values for the participants with SCI would most likely have been higher than their actual vital capacities, whereas predicted values for individuals without SCI should have been fairly close to the actual values obtained. This premise is based on an investigation by Hoit and colleagues (1990) who reported reductions in the actual vital capacities of individuals with SCI compared to their predicted values. Thus, calculation of inspiratory volume excursions as a percent of predicted vital capacity would have resulted in a greater difference between the two groups within the present study.

Just as no significant difference existed between groups for the pre-speech inspiratory variable, none existed between speech samples either. Differences in the percent of the vital capacity consumed per pre-speech inspiration were predicted to occur between dictation that involved spontaneous speech and reading aloud. This hypothesis was based on previous research (Hodge & Rochet, 1989) documenting that end-inspiratory levels (expressed as %VC relative to REL) were somewhat higher for reading than for conversational speech due to the enhanced preplanning that reading affords the respiratory system. Thus, it was hypothesized that the %VC inspired per pre-speech inspiratory gesture would be larger for reading than for spontaneous speech within the current study, but no such significant differences were found. An explanation for this finding may be found in a hypothetical premise that was too restrictive. In fact, if the results of Hodge and Rochet (1989) for both inspiratory and expiratory data are considered, end-expiratory levels expressed as %VC relative to REL also were higher for reading, which suggests that their speakers initiated and terminated the reading task at higher levels in their lung volume, but did not actually exchange a greater percentage of their vital capacity when compared to the spontaneous speech condition. Results from work completed by Hoit and Hixon (1987) confirm that %VC inspired across reading and extemporaneous speaking activities are highly similar. Hoit, Hixon, Altman and Morgan (1989) also collected speech-breathing data during extemporaneous speaking and reading tasks, and found that the results across the two speech tasks were highly similar across women in three age groups. Thus, the lack of difference between the speech samples for %VC exchanged per pre-speech inspiration in the current investigation is what should be expected based on the literature.

Mean Volume (%VC) per Expiratory Segment for Speech

This variable reflected the average amount of air expired during a respiratory cycle that supported speech, expressed as percent vital capacity. The expiratory segment volumes exhibited by the control group in the present study during dictation without ASR ranged from 13.69 to 15.71 %VC across both speech samples. This range is consistent with the low end of those reported in other studies of non-disabled individuals (Hodge & Rochet, 1989; Hoit & Hixon, 1987; Hoit, Jenks, Watson & Cleveland, 1996; Winkworth & Davis, 1997). For example, Winkworth and Davis (1997) report volumes per expiratory segments ranging from 16.04 to 19.28 %VC across reading aloud and monologue. Hoit and Hixon (1987) report lung volume excursions that ranged from 14.22 to 20.63 %VC across three age groups for reading and extemporaneous speech. Hoit, Jenks, Watson and Cleveland (1996) report lung volume excursions that ranged from 14 to 23 %VC during a normal reading activity. Hodge & Rochet (1989) report lung volume excursions that ranged from 13.4 to 22.9 %VC across tasks of reading and conversation.

Among the independent variables compared in this study, significant results for % VC expired for speech were found among dictation conditions only. Results for speaker group and speech sample were nonsignificant. Percent VC expired for speech was predicted to be lowest in the discrete-word ASR condition and highest in the control condition. This prediction was based on the same expectations that influenced the hypotheses for pre-speech inspirations, namely the effects of the constraints of ASR software on speech for successful recognition and the demands for respiratory support based on the anticipated length of breath groups.

Although differences found in %VC expired were consistent with *a priori* predictions, those differences were significant only between dictation with discrete-word ASR and the other two dictation conditions. While volume per expiratory segment for speech was less in the continuous-speech condition than in the control condition, the difference was not significant. As was the case for pre-speech inspiratory volumes, the findings suggest that discrete-word ASR constrains expiratory behavior for speech unnaturally but that continuous-speech ASR does not constrain expirations to a degree that differs significantly from control values. This is sensible in the context of the report by Winkworth and colleagues (1995) who found differences not only in the depth of inspiration at different linguistic boundaries (i.e., clauses vs. single words), but also in the volume expired for breath groups that consist of a clause or several clauses compared to that expired for shorter utterances such as single words or fillers. Because natural dictation and dictation with continuous-speech ASR are more likely to be characterized by breath groups that consist of at least one or several clauses, those speaking conditions are likely to be associated with larger expirations than those required for single-word utterances. It appears that the regulation/control parameters of discrete-word ASR required for successful dictation bring about expiratory behaviors that would be expected for single-word utterances.

A priori, it also had been hypothesized that the % VC expired for speech would be larger in individuals with SCI than in the control participants. This hypothesis was based on documented speech-breathing differences in individuals with SCI (Hixon & Putnam, 1983; Hoit, Banzett, Brown & Loring, 1990) that also pertained in the previous discussion of pre-speech inspiratory volumes. In the current investigation, although the between-group results were consistent with the experimental difference proposed [individuals with SCI expired a greater proportion (16 %VC) of their vital capacity for speech than did individuals without SCI (12% VC)], the ANOVA revealed no significant difference between groups in %VC expired for speech. That expected differences between groups for %VC expired were apparent but not significant in the present study is not surprising considering the nonsignificant results that were observed between groups for pre-speech inspiratory volumes. The depth of inspiration during speech is dictated by the alveolar pressure demanded to support an ensuing utterance (Hixon, 1973) and therefore, it is likely that the inspiratory and expiratory volumes exchanged for speech are highly correlated. Thus, it is likely that similar results between %VC inspired and expired would be found because of the nature of the respiratory system and a speaker's desire to use the available air supply for speech in a manner that requires the least respiratory muscle recruitment and force.

It was predicted that %VC expired for speech would be higher during dictation of a read passage than during dictation with spontaneous speech. This hypothesis was stated in consideration of previous research by Hodge and Rochet (1989) documenting that end-expiratory levels expressed as %VC relative to REL were slightly higher for reading than for conversational speech (Hodge & Rochet, 1989). There were no significant differences in %VC expired between speech samples in the current study. As was the situation for %VC inspired, the hypothetical premise may have been too restrictive. In fact, lack of a significant difference across speech samples for % VC expired may be what is expected based on previous research suggesting that lung volume exchange for reading aloud and speaking spontaneously are highly similar (Hodge & Rochet, 1989; Hoit & Hixon, 1987; Hoit, Hixon, Altman & Morgan, 1989). Thus, the findings for %VC expired for speech between speech samples is what should be expected based on the literature.

Average Absolute Volume Expenditure per Expiratory Segment for Speech

Data for average absolute volume expenditure per expiratory segment for speech were derived from the same data that were used to compute %VC expired for speech, expressed in milliliters (ml). Thus, this discussion of the findings for absolute volume expenditure for speech are related to the results for %VC expired for speech amongst dictation conditions, between speaker groups, and between speech samples.

Average absolute volume expenditure exhibited by the control group in the present study ranged from 517 to 599 ml during dictation in the control condition across both speech samples. Some studies have reported lower values for non-disabled individuals (Sperry & Klich, 1992; Winkworth Davis, Adams & Ellis, 1995; Winkworth, Davis, Ellis & Adams, 1994); whereas, others have reported higher values (Solomon & Hixon, 1993). Winkworth and colleagues (1994) reported a mean volume expired value of 443 ml for their female subjects during a reading task and a mean volume expired value of 441 ml during a spontaneous speech task (Winkworth et al., 1995). Sperry and Klich (1992) reported a range of 454 to 497 ml for women of two age groups across a reading task. In contrast, Solomon and Hixon (1993) reported a range from 769 to 777 ml across a monologue and reading task in a group of healthy men. The expiratory volumes reported in the present study fall between the ranges reported by other researchers. The differences between the results reported in other studies and the current study might be related to the method of respiratory data collection and the subjects who were studied.

With respect to the method of data collection, the studies that reported lower mean expiratory values (Sperry & Klich, 1992; Winkworth Davis, Adams & Ellis, 1995; Winkworth, Davis, Ellis & Adams, 1994) used a method known as respiratory inductive plethysmography (RIP) to collect respiratory data. This method involves the use of two elasticized bands, one placed around the abdomen and one placed around the chest, to transduce rib cage and abdominal wall movement. Although similar results to magnetometric data are obtained with RIP when it is used carefully, it is prone to measurement artifacts related to slippage of the bands (Hixon & Hoit, 1984; 1985). Hixon and Hoit (1984) suggest that if the respibands are placed over a subjects' clothing, then the sensors and the subject's chest wall move in an uncoupled fashion. Winkworth and colleagues (1994; 1995) placed the bands over a lycra garment and fastened them with an elastic retaining singlet, while Sperry and Klich (1992) placed the bands over a tshirt and fastened them with pins. Although this suggests that care was taken in placing the bands, it is possible that the garments the subjects wore slipped into a different position than that at the outset of the experiment. If a garment were to slip into a position on the chest wall that was characterized by smaller anterioposterior displacements than typical positioning, volume contributions may appear to be smaller than if the bands were maintained at points of maximum excursion.

The difference in mean volume per expiratory segment of speech recorded between the current investigation and that of Solomon and Hixon (1993), both of which used magnetometers, may be related to the subjects who were studied in each investigation and the method of data measurement. The subjects in the Solomon and Hixon study were male; those in the current investigation were male and female (4 males, 2 females). Sex differences in mean volume expenditure for speech have not been directly studied, but some inferences can be made from differences described in cc/syllable expired between men and women. Hoit, Hixon, Altman & Morgan (1989) report that women expend a smaller volume of air per syllable than men. Thus, for a consistent number of syllables spoken at a consistent rate, women would expend less volume than men. Therefore, because the control group in this study included two women, the mean expiratory values for speech reported herein may have been lowered by inclusion of the female data. To control for possible differences in absolute volume expenditure that may be a function of sex, %VC expired was calculated within the current investigation.

In addition to subject matters, the method of data measurement between the current study and that of Solomon and Hixon (1993) differed. Respiratory data collected by Solomon and Hixon were digitized using a digitizing pad. Although not substantiated in the literature, digital analysis of speech breathing traces is likely to be less error prone than the visual method used in the current investigation. Thus, the data reported herein may not agree with the Solomon and Hixon data because of measurement artifact. Great care was taken in the current study with respect to the voltage measurement of the expiratory traces from the oscilloscope, however, it is possible that measurement artifacts could possibly under- or overestimate volume reality and thus, account for differences between the current study and others. Measurement artifacts are likely to be inherent in the data reported across all of the speech-breathing studies, leading to natural measurement variation above and beyond any differences due to the method of data collection and subject characteristics.

Expiratory volumes were predicted to differ amongst dictation conditions with volumes being lowest when dictating with discrete-word ASR and highest when dictating in the control condition. Like the hypothesis generated for %VC expired, this prediction was based on the respiratory demands of shorter versus longer speech segments for interaction with discrete-word ASR software versus more natural speaking conditions. Although the data support the expected direction of the hypothesis, the ANOVA results reveal that differences were only significant between dictation with discrete-word ASR and the other two dictation conditions. While expiratory volume for speech was less in the continuous-speech dictation condition than in the control condition, the difference was not significant. The findings, like those revealed for %VC expired, reinforce the observation that discrete-word ASR constrains expiratory behavior for speech unnaturally, but that continuous-speech ASR does not constrain volumes expired to a degree that differs from those exchanged during more natural speech. Thus, as was the case for %VC expired, the regulating influence of discrete-word speech recognition affects the respiratory system in a manner that would be expected for short utterances.

Expiratory volumes in individuals with SCI were expected to be influenced by the same respiratory control and capacity limitations that influenced the results for these participants reported for %VC expired. Thus, it was hypothesized that absolute volume expenditure for speech would be lower in these individuals than in their non-disabled controls. However, as was the case for %VC expired, the differences in absolute volumes expired between groups was not found to be significant. Because absolute volume expired and %VC expired are based on the same data, it is likely that they are highly correlated and thus exhibit similar results. Although the ANOVA results did not reveal any significant between-group differences, the data confirm the direction of the *a priori* hypothesis: overall, individuals with SCI expended less volume per expiratory segment (343 ml) than did individuals without SCI (480 ml).

Mean volume expired for speech was predicted to be the same during dictation of a standard passage and spontaneous dictation of a letter. This hypothesis was based on research that has revealed similarities in expiratory lung volume exchanged between reading and conversational speech (Hodge & Rochet, 1989; Solomon & Hixon, 1993; Winkworth et al, 1994; Winkworth et al., 1995). The hypothesis was supported by the data collected herein and is consistent with comparable reports in the literature, suggesting that the type of speech sample chosen to elicit speech-breathing behaviors does not have an impact on the absolute volume expenditures associated with speech.

Duty Cycle of the Respiratory System within the Expiratory Phase for Speech

The duty cycle of the respiratory system for speech was a measure taken to describe the portion of expiration that was filled with speech across dictation conditions, speakers and speech samples. To the primary investigator's knowledge, measures of duration similar to those that defined duty cycle in this study do not exist in the literature. However, some attention has been paid to prephonation and postphonation air loss during an expiratory gesture that includes speech in previous research (Leeper, 1976; Sperry & Klich, 1992). As well, other authors (Winkworth, Davis, Ellis & Adams, 1994; Winkworth Davis, Adams & Ellis, 1995) have reported two different measures to represent the expiratory side of the respiratory cycle: mean volume displaced per expiratory segment and speech volume displaced per expiratory segment. These reports suggest that the expiratory side of the respiratory cycle is not always filled completely by speech. When the percentage of time that was spent in speech during expirations was explored across the experimental variables in the present study, differences were found across dictation conditions and between speech samples.

Duty cycle was predicted to differ between dictation conditions, being lowest when dictating with discrete-word ASR and highest in the control condition. This hypothesis was based on the effects of ASR software on speech that were documented during pilot-study work, in which expiratory segments in the discrete-word dictation condition were associated with short breath groups that did not fill the expiratory segment, while expiratory segments produced during dictation in the control condition were associated with speech production that appeared to fill the expiration. This observation was supported by the results of the present study; a significantly greater portion of the expiratory segment was used for speech in the control condition than in either of the ASR conditions. This suggests that the use of available respiratory support is not maximized when dictating with either type of ASR software.

Duty cycle also had been hypothesized to differ between spontaneous speech and reading of a paragraph. Specifically, it was predicted that duty cycle of the respiratory system would be higher for reading than for spontaneous speech. The hypothesis put forth for duty cycle between speech samples in the current study was supported by the data collected. A greater proportion of the expiratory segment was filled with speech when participants read than when they spontaneously dictated a letter. This outcome is also supported by the research of Solomon and Hixon (1993) in which their speakers produced more syllables and used more time for speech per expiratory segment during reading than during monologue. This difference may be related to the reduced linguistic formulation demands associated with oral reading. Pauses and other interruptions in read speech are minimal and therefore, the forward flow of speech is not broken up (Hodge & Rochet, 1989; Solomon and Hixon, 1993). It is believed that pauses and interruptions would reduce the time that is spent in speech on an expiratory segment. In fact, Solomon and Hixon (1993) found that 20% of all breath groups during a monologue task contained at least one pause with an average duration of 70 msec; whereas only 7% of breath groups during a reading task contained a pause with an average duration of 45 msec. Thus, pauses occurred more often and were longer during monologue than reading, a situation that may decrease the duty cycle for expiration when speaking spontaneously.

Duty cycle was also predicted to differ between individuals with and without SCI due to the constraints placed upon the respiratory systems of individuals with SCI.

Manning and colleagues (1992) studied individuals with and without SCI to determine differences in whole-body oxygen consumption during resistive-loaded breathing tasks. Their results suggest that respiratory muscle oxygen cost was greater in individuals with SCI. These authors hypothesized that one of the contributing factors to the differences found was related to differences in the mechanical advantages of the inspiratory muscles that were accessible for each group. For example, within the group of individuals with SCI, the muscles available for inspiration were the accessory muscles such as the trapezius, sternocleidomastoid, scalene, and pectoralis muscles. The authors speculated that the accessory muscles are placed less strategically than the intercostals to move the rib cage. Thus, the dependence of individuals with SCI on accessory muscles for inspiration may have contributed to their inefficiency relative to the group of control subjects. These investigators concluded that the decreased efficiency of inspiration in individuals with SCI could compound respiratory muscle weakness and lead to a premature onset of fatigue in situations where the inspiratory muscles are faced with a greater workload. Based on this information, it was expected that individuals with SCI in the current investigation might employ strategies to take advantage of their available expiratory supply for speech to as great an extent as possible in order to reduce respiratory workload, especially in the ASR dictation conditions.

Although the results of the ANOVA analysis revealed no significant between-group differences, the between-group data collected in the present study approached significance (p=.042) and confirmed the expected direction of differences in duty cycle between individuals with SCI and their non-disabled cohorts. The grand means for duty cycle across all dictation conditions revealed that individuals with SCI used a greater proportion of their expiratory segment for speech (69%) than did individuals without SCI (63%). Interestingly, when looking at duty cycle within each dictation condition, trends did not reveal differences between groups in the control condition (90% for individuals with SCI versus 91% for their non-disabled cohorts). However, during dictation with discrete-word ASR, individuals with SCI tended to use more of the expiratory cycle for speech (55%) than the individuals in the control group (46%). Likewise, when dictating with continuous-speech ASR, individuals with SCI tended to use more of the expiratory cycle for speech (62%) than the individuals in the control group (54%). This suggests that the individuals with SCI were more efficient with their respiratory supply when faced with the demands associated with ASR systems than the individuals who had normal respiratory function.

Average Duration of Inspiratory Pause Time

The average duration of inspiratory pause time was the time subjects took to inspire before an expiratory segment for speech. Normal inspiratory pause time for speech breathing in non-disabled individuals is approximately 0.6 seconds (Horii & Cooke, 1978). For a control group of 14 healthy men, Solomon and Hixon (1993) report inspiratory pause times that ranged from 0.588 to 0.738 seconds across samples of reading and monologue. The mean inspiratory pause times of non-disabled individuals in the present study ranged from .549 to .581 seconds during dictation in the control condition across speech samples.

When inspiratory pause times were compared across dictation conditions, speakers, and speech samples, significant differences were exhibited among dictation conditions only. It had been hypothesized that inspiratory pause time would be shortest when dictating with discrete-word ASR and longest in the control condition. This hypothesis was related to the expected outcomes for %VC exchanged per pre-speech inspiration. For that hypothesis, %VC inspired was expected to be lowest when using discrete-word ASR because of the expectancy for shorter breath groups, and highest when speaking in the control condition because of the expectancy for longer breath groups. Thus, it was reasoned that a smaller %VC inspired would be associated with shorter inspiratory pause times, and vice versa. This hypothesis was not supported by the data collected. Although %VC inspired for speech was found to be significantly larger in the control condition than in the discrete-word ASR condition, inspiratory pause time was shorter in the control condition than it was when subjects dictated with ASR. The difference in inspiratory pause time between the control condition and the two ASR conditions was significant, while that between the discrete-word and continuous-speech conditions was not significant. While the inspiratory pause time found in the control condition reflected durations within expected limits (0.57 seconds), inspiratory pause times in the ASR conditions were significantly elevated (0.84 seconds for continuous-speech and 0.79 seconds for discrete-word ASR). Thus, it appears that inspiratory pause times were inversely related to inspiratory volumes.

One explanation for this outcome may be related to the amount of cognitive processing that is associated with each type of dictation. Presumably, more processing would be required during inspiratory phases of the dictation process with either type of ASR system, because a speaker must wait to read the ASR output on the computer screen, check the accuracy of the printed message, and decide what command to give the computer in the case that the message has not been understood. As suggested by Hodge and Rochet (1989), situations that are characterized by higher linguistic formulation demands may require a greater amount of a speaker's attention, resulting in speech-breathing performance that is altered from that supporting speech under less demanding conditions. Thus, it appears that the regulating effects of the ASR systems were determinants of both inspiratory pause time and the volume of air that needed to be inspired for speech.

With respect to speaker-group differences for this variable, inspiratory pause time was hypothesized to be longer for individuals with SCI than for those without. This hypothesis was based on a case study by Hixon and Putnam (1983) in which an individual with cervical spinal cord injury exhibited inspiratory pause times that were longer than those reported for individuals without SCI. The results of the present study, however, did not reveal such a difference for the speakers with SCI. Their mean inspiratory pause time across dictation conditions and speech samples was 734 msec (SD = 190 msec), and that for the control subjects was 736 msec (SD = 160 msec). The lack of a significant difference between groups is curious in light of the impairment of the respiratory muscles in individuals with SCI. For example, subjects CM and EJ had C6-7 lesions, which would indicate that diaphragmatic function was probably intact. Other subjects, PM, MH and JA had C5-6 lesions and GC had a C4-5 lesion, which may indicate that innervation to the diaphragm was at least partially intact. The levels of lesion across these individuals also suggest, however, that intercostal and abdominal muscles were paretic for all, which implies weakness with the potential to affect inspiratory pause times. As stated by Hixon and Putnam (1983), slow inspirations are related to two factors: paresis of the inspiratory muscles of the rib cage and diaphragm, and paresis of the abdominal muscles. The loss of abdominal muscle function would impair a speaker's ability to stabilize the abdominal wall during inspiration thereby decreasing the mechanical advantage for contraction of the diaphragm. Thus, it was anticipated that the inspiratory pause time for individuals with SCI would be longer than that for individuals in the control group because of the abdominal paresis associated with the SCI and its influence on inspiratory efficiency. However, the current results do not support this hypothesis.

The remaining hypothesis about inspiratory pause times expected them to be shorter for reading than for conversational speech. This was based on research that reported this pattern of pre-speech inspiratory pause time across speech samples in a group of healthy men (Solomon & Hixon, 1993). In addition, the hypothesis was based on research suggesting that, because reading follows a script, it requires minimal linguistic formulation (Hodge & Rochet, 1989). Therefore, reading may be a less-demanding cognitive task than spontaneous speech because the speaker does not need extra time between breath groups to formulate linguistic content. Research that supports this notion was reported by Mitchell, Hoit and Watson (1996) in which inspiratory durations were found to be longer for an extemporaneous expressive linguistic task that was produced without an outline (a presumably harder task than one in which an outline was given).

Notwithstanding expectations based on research that has reported differences in inspiratory pause time between tasks of reading aloud and spontaneous speech, outcomes for this variable in the present study showed no such differences.

One final consideration for differences in pause time may be related to the loudness level at which speakers were operating. In another study of respiratory behavior across different speaking conditions, a relationship has been found between the intensity, or loudness, of an utterance and the amount of air inspired in a particular frame of time. Russell, Cerny & Stathopoulos (1998) reported that greater volumes of air were inhaled in shorter amounts of time when their subjects spoke with louder voices. Thus, it is possible that changes in loudness across the experimental conditions in the present study may have a bearing on the results found for inspiratory pause time and %VC inspired.

Speech-Breathing Data Concluding Remarks

For the speech-breathing variables in this investigation, differences amongst dictation conditions were most obvious. Trends towards differences were observed between speaker groups, but no significant differences were found. Significant differences were found between speech samples for one dependent variable only.

Observation of the differences among dictation conditions suggests that discrete-word ASR perturbs the speech-breathing mechanism to the greatest degree, influencing inspiratory and expiratory volumes (%VC) and mean volume expenditure (ml) to a greater degree than the control condition or continuous-speech ASR. For the same dependent variables, continuous-speech ASR did not constrain the respiratory system to a degree that was significantly different from natural (control) dictation. This was not the pattern observed for duty cycle or inspiratory pause time, however. For these variables, both ASR systems perturbed the speech-breathing system in a manner that was different from natural speech. Another way to consider the speech-breathing variables across dictation conditions is that behaviors related to volume exchange did not appear to be constrained unnaturally by the regulating effect of continuous-speech ASR, whereas data related to temporal variables did appear to be constrained unnaturally by continuous-speech ASR. In contrast, speech-breathing variables related to *both* volume and time appeared to be constrained unnaturally by discrete-word ASR.

These volume and time issues can be considered in terms of workload, efficiency, and fatigue. As per the definition put forth in Chapter 1, an increased workload related to the speaking process will produce a measurable change in that process. Changes in the volume-related variables measured during the use of ASR in this study imply a potentially higher workload for a speaker, especially while using discrete-word technology. The decrease in the available volume of air for speech while dictating with discrete-word ASR implies that fewer words will be spoken on one expiration, which will ultimately lead to an increase in the number of respiratory cycles needed to dictate a message. Thus, the potential for increasing the speech-production workload is imminent, which is a concept that will be explored further in Chapter 5. With respect to efficiency, the variables that most clearly address this concept are those related to time. The duty cycle of the respiratory system within the expiratory phase for speech was significantly lower when individuals dictated with ASR software suggesting that dictation with ASR is not as efficient in its use of the expiratory breath supply as natural dictation. A similar pattern of behavior was observed for inspiratory pause time. Inspirations when dictating in the control condition were quick and efficient, and supported normal expiratory volumes that were larger than those seen for dictation with ASR. Dictation with both ASR systems was characterized by significantly slower inspiratory pause times than control dictation, and these inspirations preceded smaller-than-control expirations for speech, especially for discrete-word dictation. The slow inspirations have been interpreted as a function of the user-ASR interface, which may be influenced by higher cognitive workload and regulation-control factors required for successful recognition with these systems. When the workload and efficiency issues related to dictation with ASR are considered together, it becomes apparent that the potential for fatigue of the speech-production system exists. This consideration will become even more apparent after the speech-production variables, which are highly dependent on the speech-breathing variables explored in this chapter, are explored in Chapter 5.

Unlike the results observed across dictation conditions, significant differences in the speech-breathing variables between speaker groups were not found. Trends in the direction of the expected outcomes were apparent between speaker groups for 4 of the 5 dependent variables (mean volume per pre-speech inspiration, mean volume per expiratory segment for speech, average absolute volume expenditure per expiratory segment for speech, and duty cycle of the respiratory system within the expiratory phase for speech), however. Thus, one of the primary concerns with respect to lack of a significant finding between groups was that the prospective power analysis value may have been set too low (power = .70), thereby underestimating an adequate sample size. With a larger sample size, it may be possible that significant differences would emerge between individuals with and without SCI.

Significant differences were revealed across speech samples for only one speech-breathing variable – duty cycle. The significant finding for duty cycle within the expiratory phase for speech suggests that reading affords the respiratory system a greater degree of efficiency than spontaneous speech because of the lower cognitive workload associated with reading. This was interpreted in light of reports (Hodge & Rochet, 1989; Solomon & Hixon, 1993) that the linguistic formulation demands associated with spontaneous speech may result in pauses and interruptions to the natural forward flow of speech, thus reducing the actual amount of speech produced on one expiration.

In conclusion, the most striking results are related to the observation that ASR systems have the potential to induce changes in the inspiratory and expiratory behavior of the respiratory system, and the degree to which the available air supply is used for speech. The structure of the linguistic message that is appropriate for either type of ASR system (i.e., single word input with pauses between each word for discrete-word ASR and connected speech input for continuous-speech ASR) appears to influence breathing behaviors to a degree that differs from speech in a natural control situation. Discrete-word ASR differed significantly from a natural speaking situation across all speech-breathing variables studied, whereas continuous-speech differed from natural speech for only three of the five speech-breathing variables that were studied. This suggests that continuous-speech ASR technology may have been successful in increasing the degree to which users can dictate in a more natural manner. It will be valuable to consider these results in the context of the kinematic and speech-production variables also studied in this research.

CHAPTER 4 - Respiratory Kinematics Results

Purpose

The intent of this portion of the study was to describe the relative chest wall part contributions to lung volume exchange associated with spontaneous speech and reading aloud in three dictation conditions (with discrete-word ASR, with continuous-speech ASR, and with no ASR) in two groups of users (those with and without spinal cord injury). In addition, the background chest wall configuration for each subject was described across dictation conditions and between speech samples.

Variables

The characteristics for this description were:

1. Rib cage contribution to lung volume exchange during dictation of a paragraph read aloud and time-limited passage of spontaneous speech (%)
2. Initiation and termination of speech relative to REL, and range of lung volume excursions
3. Background configuration of the chest wall during dictation of a paragraph read aloud and time-limited passage of spontaneous speech

Method

Data Collection, Processing and Analysis

Formatting and Calibration

A two-channel digital storage oscilloscope with x-y display capability was used for the estimation of relative chest wall part contributions to lung volume exchange and associated background chest wall configuration. Signals from the rib cage (RC) and abdominal (ABD) magnetometers that were stored on FM tape were sent to separate channels of the oscilloscope such that the ABD influenced the x-amplifier and the RC influenced the y-amplifier; the display represented a relative diameter chart (See Figure

2-2). In the x-y display mode, the experimenter began the kinematic analysis of each speaker's data by adjusting the deflection of the beam for the isovolume maneuver at REL using the output gain controls on the FM tape recorder so that the slope of the isovolume line was -1 (135°). Kinematic theory (Hixon et al., 1973) states that, during the isovolume maneuver, the volume exchanged between the RC and ABD is equal and opposite. Thus adjusting the signal to a slope of -1 during a participant's performance of an isovolume maneuver allows an estimate of the relative volume exchange of the RC and ABD to be made from that person's subsequent kinematic tracings recorded during speech. Some of the isovolume lines obtained from subjects in this study were characterized by hysteresis, which is not unusual for maneuvers performed by naïve subjects (Hixon, Goldman & Mead, 1973; Hodge & Rochet, 1989). When this occurred, the tracing that corresponded to inward abdominal movement was used to represent the isovolume line. The decision to use the position of the isovolume trace representing inward movement of the abdominal wall was based on the premise that this would be the movement most likely observed during abdominal muscular contribution to speech for non-disabled individuals. For all subjects with SCI in this experiment, this movement reflected the action of the primary investigator, who assisted with the displacement of volume by gently pushing and then releasing the abdominal wall. In addition, not all isovolume lines were perfectly straight; some were characterized by curvilinearity at the extremes of the maneuver. Therefore, the segment of the isovolume line that exhibited a relatively constant straight portion was adjusted to a slope of -1.

Once the slope adjustment had been made, a line representing the isovolume maneuver displacement at REL was drawn on top of the fluorescent trace on the oscilloscope screen in non-permanent ink. For marking the second isovolume maneuver at a specified lung volume above REL, the experimenter allowed the FM tape to run and observed a subject's quiet tidal breathing patterns on the oscilloscope. At the end of this period, when cues from the banter channel (channel 4) of the FM tape indicated that the subject was preparing for the next isovolume maneuver, the experimenter adjusted the beam on the oscilloscope screen when the subject expired to REL (before inhaling on the respirometer) so that the point of end-expiration touched the REL isovolume line. Then, as the FM tape continued to play, the subject's inhalation to a specified lung volume and performance of an isovolume maneuver at that lung volume were captured on the oscilloscope screen. The track of the oscilloscope beam for this second isovolume maneuver also was reproduced on the screen in ink. The actual volume inspired and level at which this second maneuver was accomplished relative to REL had been announced on tape by the investigator during the study and verified by reference to the inhalatory tracing on the respirometer recording paper. Once the isovolume reference lines had been captured and marked, periods of resting tidal breathing from the tape record were observed and their excursions marked on the oscilloscope screen relative to the isovolume lines. The beam was adjusted if necessary so that at tidal end-expiratory moments it came to rest on the REL isovolume line that was traced on the oscilloscope screen. Once the experimenter was confident that resting tidal volume was stable relative to the isovolume line at REL, the magnetometric data associated with speech dictated in each condition were played so that speech breathing traces could be captured on the oscilloscope relative to the isovolume and the tidal breathing lines.

For the California Passage, kinematic data for 45 seconds of recorded speech were captured on the oscilloscope from the start of the phrase, "…, *the mountains of California are favorite vacation spots*". A line was traced through the middle of the pattern that accumulated on the screen, which represented the history of RC and ABD diameter changes across the 45 seconds of the passage read in each dictation condition. For spontaneous dictation of a letter, data were collected on the oscilloscope beginning at 45 seconds into a speaker's monologue and continuing through the next 45 seconds of recorded speech. The examiner traced through the middle of the pattern of these breath groups in the same manner as she had documented the chest wall kinematic traces for the California Passage. For reliability purposes, the investigator repeated the isovolume markings made, and recaptured the resting tidal breathing and dictation traces in one condition chosen randomly for each subject.

Once all markings associated with a particular series of tidal breaths or speech breath groups had been made on the oscilloscope screen, tracing paper was placed overtop them and their patterns were reproduced on the paper along with the isovolume and tidal reference lines. The tracings were then digitized using a Dexxa color scanner (4800, Storm Technology, Palo Alto, CA). The scanned images were transferred into Adobe Photoshop (Version 5.0, Adobe Systems, Inc., San Jose, CA) and saved as .jpeg files. These files were then imported into Microsoft PowerPoint (97), and the drawing tools in that program were used to reproduce the images included in this chapter.

Kinematic data for 11 of the 12 subjects are described in this chapter. Kinematic data for MH were excluded from this descriptive analysis. His magnetometric data on

FM tape were too unstable to provide reliable baseline information required for the assessment of background chest wall configuration.

Relative Chest Wall Part Contribution To Lung Volume Exchanged

Rib cage contribution to volume exchanged for both tidal breathing and speech was estimated through measurement of the x-y kinematic data displays. The slope of the traced pathway representing approximately 30 seconds of tidal breathing before initiation of the speaking tasks was measured to estimate the percent of rib cage (%RC) contribution to volume exchanged across tidal breaths. The slopes of the traced pathways representing 45 seconds of speech while reading and while speaking spontaneously were measured to estimate the %RC contribution to volume exchanged across breath groups during each speaking task. The %RC contribution to lung volume exchange was determined for the most extensive portion of each expiratory pathway that exhibited a constant slope. A manual method was used in this process (Hodge & Rochet, 1989; Putnam & Hixon, 1984). The horizontal base (zero line) of a protractor was aligned in parallel with what represented the x-axis on the oscilloscope screen. The centre point of the protractor was placed on the vertex of the angle created by the intersection of the speech-breathing trace with the isovolume line at REL, and the angle of the speech breathing trace was then determined from degree markings on the protractor. Percent RC contribution was calculated by dividing the degree of the slope of the speech breathing line by 90°, and then multiplying by 100. This method was used to estimate percent rib cage contribution during both tidal and speech breathing for each of the participants.

Initiation and Termination of Speech Relative to REL & Range of Lung Volume Excursion

Interpretation of initiation and termination of speech relative to REL was completed by determining whether a speech-breathing trace was initiated or terminated at a level greater than one litre above REL, between REL and 1 litre above REL, or below REL. Results were tallied for dictation in each condition for individuals in both speaker groups. Interpretation of the extent of lung volume excursions was completed by visually examining the length of the speech-breathing lines for each individual across dictation conditions and between speech samples. For ease of interpretation of initiation and termination levels relative to REL and range of volume expenditure, the primary investigator adjusted the isovolume lines above REL to 1 litre on a working copy of the kinematic charts for each subject. When this was done, the extent of the volume expenditures could be examined for each individual with reference to that line.

Background Configuration of the Chest Wall

Chest wall relaxation curves were generated by the subjects in this experiment through relaxation maneuvers at three lung volume levels (near total lung capacity, near resting expiratory level, and a point roughly halfway between the two). Additionally, each subject produced a series of three sighs, which were meant to provide an approximation of the relaxation curve. A review of the FM-recorded kinematic tracings revealed that many of the curves derived from the relaxation maneuvers did not approximate the curves derived from the sighs, however. This is likely due to the complexity of the task of generating a relaxation curve. It has been reported that subjects often lack the sophistication necessary to reliably produce such curves within the time-restraints of many experiments (Hodge & Rochet, 1989). Thus, the relaxation curves produced by the subjects in this experiment were considered unreliable as reference points for assessing the placement of other kinematic data. Although reference points for kinematic data on relative volume charts are traditionally related to an imaginary relaxation curve when a tangible reference is unreliable, chest wall configuration during speech can be inferred from kinematic data if information is available about chest wall configuration during resting tidal breathing (Hodge & Rochet, 1989; Putnam & Hixon, 1984). This alternative was suggested by Putnam and Hixon (1984) who reasoned that "tidal end-expiratory level…usually corresponds to the resting level (FRC) of the respiratory apparatus, and the point at which the tidal expiratory pathway intersects the FRC isovolume line most nearly approaches the relaxed configuration of the chest wall at that level" (p.303). Therefore, the point at which the expiratory tidal breathing lines generated by the subjects in this experiment intersected the REL (REL=FRC in the preceding quote) isovolume line were used as an approximation of a point of relaxed configuration of the chest wall.

With a reference point of relaxation at REL determined, interpretation of the configuration of the chest wall system can be related to this point and the tidal breathing line associated with it. As stated by Hixon and colleagues (1973), movement up (i.e., to the left on) any isovolume line involves a shifting of volume from the abdomen to the rib cage; and conversely, movement down (i.e., to the right on) any isovolume line involves a shifting of volume from the rib cage to the abdomen. Thus, with a relaxation line of reference in mind, data that lie within the isovolume contours to the left of the tidal breathing trace reveal rib cage diameters that are larger and abdomen diameters that are smaller than those found closer to relaxation. The opposite is true for data that lie to the right of the presumed relaxation state. Each pattern of speech-breathing traces obtained for each subject in each dictation condition was judged as to whether it fell to the left of, right of, or coincident with a subject's resting tidal breathing trace for information about background torso configuration. Speech-breathing lines that fell to the left of the resting breathing line indicated a torso configuration in which the rib cage was larger and the abdomen was smaller (i.e., an inverted pear shape) than their relaxed size at the same lung volume. The converse was true for speech-breathing lines that fell to the right of the resting breathing line.

Figures 4-2 through 4-16 represent the respiratory kinematic data collected for the individuals in this study and traced from the digital storage oscilloscope in the form of relative volume charts. In these charts, volume displacement of the RC increases upward along the y-axis. Volume displacement of the ABD increases rightward along the x-axis. The solid diagonal line marked REL represents the adjusted slope for an isovolume maneuver at REL for each subject. The upper diagonal line represents the adjusted slope for an isovolume maneuver at a specified lung volume above REL. The dark bold dotted and solid lines represent tracings obtained during dictation of a letter and the California Passage, respectively, and the thin dotted black lines represent tidal breathing.

Descriptive Reports

Percent RC contribution data are presented in a descriptive manner in the tables and figures to follow. Initiation and termination levels relative to REL for each group are also presented in a descriptive table format, and extent of lung volume excursions between groups, dictation conditions and speech samples are estimated within the text of this chapter. Finally, the kinematic data in the figures are interpreted for information about the background configuration of the chest wall for each subject across dictation conditions and between speech samples. The measurements on which these descriptions are based were performed independently by the primary investigator and a second person with experience in the interpretation of relative volume displays of chest wall kinematic data. Intrajudge and interjudge reliability were analyzed via intra-class correlations (ICC) for 100% of the %RC contribution data collected.

Results and Discussion

Relative Chest Wall Part Contribution to Lung Volume Exchange

Percent Rib Cage Contribution

Table 4-1 contains vital capacity, isovolume slope, and % RC contribution measurements for each subject during tidal breathing and during 45 seconds of reading aloud from the California Passage and 45 seconds of spontaneous dictation of a letter in all three dictation conditions.

				%Contribution RC					
				California			**Letter**		
Subject	**Vital Capacity (litres)**	**Isovolume Slope (°)**	**Tidal Breathing (% RC)**	**no ASR**	**continuous-speech ASR**	**discrete-word ASR**	**no ASR**	**continuous-speech ASR**	**discrete-word ASR**
PM*	2.3	133	72	91	100	100	96	100	78
KT	3.7	130	76	83	92	88	89	97	89
CM*	2.3	140	62	89	91	89	62	92	75
FM	3.5	140	100	100	97	100	100	100	100
JA*	2.3	135	76	77	87	70	77	91	76
BS	3.3	140	17	13	22	44	11	42	26
GC*	1.7	133	78	100	96	100	100	93	94
RT	4.4	136	67	89	86	80	81	92	76
EJ*	3.2	132	98	88	89	89	89	96	83
GG	5.8	130	68	50	67	61	56	56	67
TA	4.3	136	71	56	56	67	61	61	58
Mean	3.4	135	71.4	76.0	80.3	80.7	74.7	83.6	74.7
S.D.	1.2	3.8	21.7	25.2	23.4	18.3	26.4	20.4	22.1

Table 4- 1. Isovolume slope and % RC contribution for tidal breathing and speaking tasks. For the speaking tasks, %RC contribution is presented for both speech samples and across all three dictation conditions. Individuals with SCI are paired with their age-, height- and sex-matched control (data for MH*, who was matched by TA, are missing). *=participant with SCI.

Isovolume Slopes

As shown in table 4-1, the slopes of the isovolume lines at REL ranged from 130° to 140° (M=135°; SD=3.8°), compared to a target slope of 135°. The primary investigator's reliability in adjusting the isovolume lines to a target slope of 135° on the oscilloscope from the FM tape record of each subject was assessed by repeating the isovolume line adjustment at least 24 hours after the first adjustment had been made. Differences between the slopes of the two tracings across all subjects ranged between 0° and 7° (M=2.3°).

Tidal Breathing - % RC Contribution

Illustration of tidal-breathing slopes can be observed in Figures 4-2 through 4-16. As shown in Table 4-1, participants' tidal breathing slopes ranged from 17 %RC to 100 %RC contribution (M=71.4; SD=21.7). Six of the 11 subjects had slopes in excess of the mean and 5 subjects had slopes below the mean. The measurement error between estimates of the slopes of the tidal breathing lines made by the primary investigator and a second person ranged from 0% to 7% across subjects, with a mean of 2%.

Speech Breathing - % RC Contribution

Illustration of speech-breathing slopes for California Passage can be observed as the bold lines in Figures 4-2 through 4-16. As shown in Table 4-1, slopes for speech breathing data on the kinematic displays ranged from 13 %RC to 100 %RC contribution (M=79.0; SD=22.3) during dictation of the California Passage. In the no-ASR condition, 7 of the 11 subjects had slopes in excess of the mean and 4 had slopes below the mean. While dictating in the continuous-speech ASR condition, 8 of the 11 subjects had slopes in excess of the mean and 3 had slopes below the mean. Finally, during dictation with discrete-word ASR, 6 of the 11 subjects had slopes in excess of the mean and 5 had slopes below the mean. The measurement error for estimates of the slopes of the speech breathing lines made by the primary investigator and a second investigator ranged from 0% to 10% across subjects in the no-ASR condition, with a mean of 2.27%; 0% to 7% across subjects in the continuous-speech ASR condition, with a mean of 1.82%; and 0% to 8% across subjects in the discrete-word ASR condition, with a mean of 2.2%.

Illustration of speech-breathing slopes for dictation of a letter can be observed as the bold dotted lines in Figures 4-2 through 4-16. As shown in Table 4-1, participants' kinematic data slopes during dictation of a letter ranged from 11 %RC to 100%RC contribution (M=77.7; SD=22.2). In the no-ASR condition, 7 of the 11 subjects had slopes in excess of the mean and 4 had slopes below the mean. While dictating with continuous-speech ASR, 8 of the 11 subjects had slopes in excess of the mean and 3 had slopes below the mean. Finally, during dictation with discrete-word ASR, 8 of the 11 subjects had slopes in excess of the mean and 3 had slopes below the mean. The measurement error for estimates of the slopes of the speech breathing lines made by the primary investigator and a second investigator ranged from: 0% to 9% across subjects in the no-ASR condition, with a mean of 3.55%; 0% to 18% across subjects in the continuous-speech ASR condition, with a mean of 4.1%; and, 0% to 14% across subjects in the discrete-word ASR condition, with a mean of 4.2%.

Differences in %RC contribution to lung volume exchange were most apparent across dictation conditions, although the differences were minimal, and appeared to differ little between speaker groups and speech samples. The similarity in %RC contribution values between speaker groups is interesting and will be considered first. Exploration of the results observed amongst dictation conditions and between speech samples will follow.

As has been found in other studies of speech breathing (Hodge & Rochet, 1989; Hoit et al, 1996), speech in the present investigation was accomplished in most instances with rib cage contribution to lung volume change being in predominance. As stated by others (Hixon, Goldman & Mead, 1973; Hixon, Mead & Goldman, 1976; Hoit, 1995; Hoit et al., 1996), predominance of the rib cage as a power source for lung volume change may be preferred for normal speakers because the rib cage covers a larger surface of the pulmonary system than the abdomen, therefore being a more efficient producer of lung volume change than the abdomen. Typically, when the rib cage is the predominant source for lung volume exchange, the abdomen remains relatively stable and contributes little to lung volume exchange above REL. In individuals with spinal cord injury, rib cage contribution may be less than what is observed in non-disabled individuals due to paresis of the intercostal muscles and a resulting inability to expand the rib cage to the same degree as non-disabled individuals. In addition, the abdomen of individuals with high-level SCI may be relatively unstable and unable to participate in lung volume exchange because of abdominal muscle paresis. However, a large range of abdominal displacement may be exhibited on the kinematic graphs of individuals with SCI, which may result because abdominal wall compliance is higher than rib cage compliance for these speakers. Small rib cage contributions and large abdominal ranges of displacement have been reported in another study of kinematic behaviors in individuals with SCI (Hoit et al., 1990).

Hoit, Banzett, Brown & Loring (1990) found that individuals with SCI tended to exhibit small rib cage contributions to volume change, as compared to non-disabled individuals. This was not a predominant observation in the current investigation. The ranges of slopes reflecting rib cage contribution for tidal breathing and speech in individuals with SCI were similar to the ranges recorded for their non-disabled cohorts within this study and to the ranges reported by others for non-disabled individuals (Hodge & Rochet, 1989). The fact that the kinematic data from the individuals with SCI in the current study do not reflect the results reported by Hoit and colleagues (1990) is intriguing. The differences between the current data and those of Hoit et al. cannot be attributed to level of lesion, as this did not seem to differ between the participants in that study and those with SCI in the current one. One explanation that may partially account for the differences between the two studies in RC contribution may be related to the time lapsed since the spinal cord injury. Half of the subjects in the study by Hoit and colleagues had relatively recent injuries when they participated in that research (less than one year post onset). The occurrence of the spinal cord injuries for the participants in the current study ranged from 2 to 28 years prior to data collection (mean = 16 years) and therefore did not include recent injuries. It has been shown by others (Loveridge, Sanii & Dubo, 1992) that there are adjustments in respiratory function up to 1 year post-SCI. Therefore, it is possible that individuals with SCI learn to compensate for their loss of respiratory muscle function and perhaps become more adept at recruiting accessory rib cage and neck musculature with time. Spared rib cage muscles, such as the pectoralis major and latissimus dorsi, and neck muscles, such as the sternocleidomastoid, scalenes and subclavius, may have been recruited more adeptly by the individuals with SCI in the current study. This may account for the fact that lung volume exchanges in this group of individuals were characterized by higher rib cage contributions.

Another explanation for the differences between the current study and that completed by Hoit et al. (1990) may be related to rib cage compliance. In the presence of a spinal cord injury, the rib cage may become less compliant (Estenne & De Troyer, 1986) due to a reduction in lung tissue elasticity and a decrease in the viscoelasticity of chest wall structures because of chronic limitations in respiratory excursions (Krachman & Criner, 1998). On the other hand, the abdomen may become more compliant due to paresis of abdominal musculature (Estenne & De Troyer, 1986). It is interesting to note that in the Hoit study, some of the individuals with long-term spinal cord injuries exhibited greater abdominal range of displacement, suggesting that they had higher abdominal compliance than rib cage compliance. If abdominal compliance is high, the rib cage will have an inadequate base against which to pressurize the respiratory system. Hixon and Weismer (1995) state that "when the abdominal wall is impaired, the rib cage wall effectively 'drops'" (p. 53). This places the rib cage muscles at a mechanical disadvantage where they will be forced to operate at lower lung volumes and will be shortened appreciably from the length assumed if they were generating normal expiratory force (Hixon & Weismer, 1995; Hixon & Putnam, 1983). Theoretically, this situation could restrict rib cage contribution to lung volume exchange. If the subjects with SCI in the current study were particularly adept at recruiting rib cage musculature to overcome such restrictions for volume change, they may have inadvertently created a configuration change. For example, Hixon and Weismer (1995) state that when compensatory activation of rib cage musculature is used to lift the rib cage to a larger lung volume, the movement may drag an impaired abdominal wall inward thereby creating a configuration change, but not necessarily an alveolar pressure change. Such configuration changes in the subjects with SCI in this study would explain why there was a lack of a large range of abdominal displacement in their kinematic data.

When these findings are considered with respect to the speech-breathing data presented in Chapter 3, some interesting points can be made. The results from Chapter 3 suggested that there were trends for differences in lung volume exchanges for speech between individuals with SCI and their non-disabled counterparts, but that these were not significant. Such trends were reflected in the kinematic data within this chapter, and like the results in Chapter 3, the differences did not appear to be great. That the individuals with SCI generated expiratory volumes for speech that were comparable to normal suggests that they recruited accessory muscles of the rib cage and neck to expand the rib cage and inflate the lungs to a degree that allowed for comparable breath group exchanges to their non-disabled counterparts. That accessory muscles could have been recruited for this purpose was exemplified in the background configuration of their chest walls, an issue that will be addressed shortly.

The individuals with SCI in the current investigation exhibited a percentage of rib cage contribution for speech that did not differ from their non-disabled cohorts and therefore, examination of differences in rib cage contributory behavior across dictation conditions will be described for all individuals. Compilation of the mean %RC contribution data across dictation conditions for both speech tasks is illustrated in Figure 4-1. Individual differences may be observed in the relative volume charts for each subject and are illustrated in Figures 4-2 through 4-16. Observation of Figure 4-1 reveals that %RC contribution was slightly higher for dictation with continuous-speech ASR during spontaneous dictation of a letter than it was during dictation with discrete-word ASR or dictation in the control condition. For dictation of the California Passage, %RC contribution was lower when dictating in the control condition, and higher when dictating with either one of the ASR systems. Thus, it appears that dictation with continuous-speech ASR in the current study tended to lead to speaking patterns in which the rib cage contributed to lung volume exchange to a greater extent than in the control-speaking situation. The effect of dictation with discrete-word ASR on %RC contribution was less clear. It appears that discrete-word ASR encouraged more rib cage contribution than natural speech when speakers dictated a scripted passage, but did not encourage such differences when they dictated spontaneously. The small differences that have been noted across dictation conditions must be interpreted with caution, however. Some of the differences, especially those observed for dictation of the California Passage, are within the measurement error for slope.

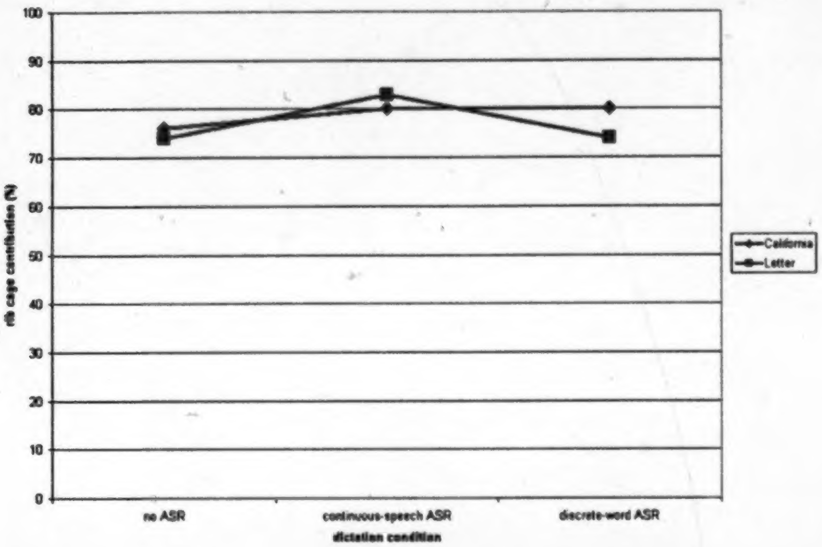

Figure 4- 1. Percent rib cage contribution across dictation conditions for all subjects for dictation of the California Passage and spontaneous dictation of a letter.

Finally, as has been found by others (Hodge & Rochet, 1989), %RC contribution did not differ between tasks of speaking spontaneously and reading. Hodge and Rochet (1989) report that %RC contribution during conversational speech was 79%, and that for reading was 81%. In the current study, mean %RC contribution for spontaneous dictation was 78%, and that for reading was 80%.

Measurement Reliability

Tidal Breathing Data Slopes

Measurement reliability was computed via intraclass coefficients (ICCs) for 100% of the %RC contribution data related to tidal breathing that were collected in this portion of the investigation. The ICC coefficient for intrajudge reliability for the primary investigator was .9977. A confidence interval of 95% was established with a lower bound of .9888 and an upper bound of .9991 within which the ICC coefficient for reliability falls. The ICC coefficient for interjudge measurement reliability between the primary investigator and another investigator experienced in the collection and measurement of kinematic data was .9957. A confidence interval of 95% was established with a lower bound of .9841 and an upper bound of .9988 within which the ICC coefficient for reliability falls.

Speech Breathing Data Slopes

Intra- and interjudge reliability were computed via intraclass coefficients (ICCs) for 100% of the %RC contribution data related to speech breathing that were collected in this portion of the investigation. The ICC coefficient for intra-examiner measurement reliability for the primary investigator was .9936. A confidence interval of 95% was established with a lower bound of .9895 and an upper bound of .9961 within which the ICC coefficient for reliability falls. The ICC coefficient for interjudge measurement reliability between the primary investigator and another investigator experienced in the collection and measurement of kinematic data was .9882. A confidence interval of 95% was established with a lower bound of .9806 and an upper bound of .9929 within which the ICC coefficient for reliability falls.

Unusual Speech-Breathing Patterns

Although the slopes of the speech-breathing traces reflect what would be considered normal %RC contribution to lung volume exchange, there were unusual patterns in the configuration of the speech-breathing traces for some individuals with SCI. Looped speech-breathing patterns were exhibited in 3 of the 5 subjects with SCI, a pattern that was not exhibited by any of the control subjects. This pattern is unusual in that it has not been reported as typical in other studies of non-disabled individuals (Hodge & Rochet, 1989; Hoit & Hixon, 1990; Hoit et al., 1996).

Looped speech-breathing patterns similar to those seen within this study have been described by Hoit and colleagues (1990) for a group of men with SCI, and by Putnam and Hixon (1983) for a group of men with motor neuron disease. In the current study, subjects PM (Figures 4-3 through 4-5) and GC (Figures 4-12 through 4-14) exhibited looped speech-breathing pathways that followed a counterclockwise course across all dictation conditions. Subject JA (Figure 4-10) exhibited looped speech-breathing pathways that followed a clockwise course and did so only in the control speaking condition. An example of a speech-breathing loop that follows a counterclockwise course for one of the speakers with SCI is illustrated in Figure 4-2.

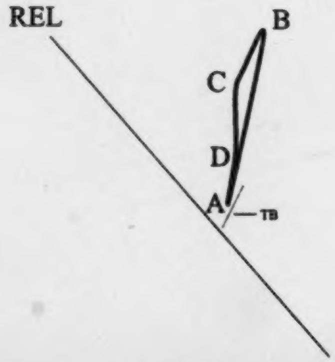

Figure 4- 2. Sample speech-breathing trace from subject with SCI (PM). Line AB denotes inspiration. Line BC denotes a sudden quiet expiratory gesture. Line CDA denotes expiration for speech. The line labeled REL represents the tracing of an isovolume maneuver at REL. The dashed line labeled 'TB' is the reference tidal-breathing trace.

Observation of Figure 4-2 reveals that the inspiratory side of the loop (Line AB) reflects chest wall part contributions characteristic of inspiration, with both the rib cage and abdomen contributing to inspiratory volume change, but with the rib cage predominating (85% RC contribution). The deformation of the chest wall along this pathway suggests that this individual with SCI could have been using accessory muscles, such as the sternocleidomastoids, the scalenes, and the trapezii, to assist with inspiration (Hixon, Putnam & Sharp, 1983). The beginning of the expiratory side of the loop is characterized initially by contributions of both the rib cage and abdomen in the expiratory direction (Line BC) (72% RC contribution), which supported a quiet expiratory gesture. Because the decrease in the size of the abdomen characterized by Line BC cannot result from active abdominal muscular force, it is suggested that this subject may have shifted her torso by leaning forward slightly and compressing the abdomen. This was followed by continued movement of the rib cage in the expiratory direction (Line CD) (100% RC contribution) and finally movement of both the abdomen and rib cage in the expiratory direction (Line DA) (85% RC contribution), both of which supported speech. It is likely that this individual used accessory muscles to check recoil forces during speech along the CDA pathway due to the fact that the trace remains to the left of the presumed relaxation characteristic.

When considering the unusual patterns with respect to dictation condition, 2 of the subjects with SCI (PM & GC) exhibited the looped configurations that occurred across all dictation conditions whereas, the other subject with SCI (JA) exhibited the looped configurations only during dictation in the control condition. For subject JA, dictation with both ASR systems was characterized by speech-breathing traces that followed the same pathway for inspiration and expiration.

Speech-breathing differences in chest wall kinematic behaviors between speech samples have been reported by Putnam and Hixon (1983) for individuals with motor neuron disease in which unusual slope changes characterized some of the subjects' conversational data, but not their reading data. Within the current investigation, unusual slope changes were observed at the top of some of the looped configurations exemplified in the speech-breathing traces of some individuals with SCI (See Figures 4-3 and 4-12). However, the kinematic data would indicate that the unusual slope changes were apparent for both scripted and spontaneous dictation. Thus, the differences between the two speaking tasks, such as the foreknowledge of linguistic boundaries in scripted speech that is not available during spontaneous-speech, did not appear to elicit distinctive breathing patterns.

Initiation and Termination of Speech Relative to REL

The kinematic data illustrated in Figures 4-3 through 4-17 were used to derive information related to initiation and termination levels of speech relative to REL. These levels were examined between groups, amongst dictation conditions, and between speech samples and will be discussed in that order.

In the control group, each individual exhibited speech-breathing traces that extended beyond REL for at least one of the dictation conditions, indicating that positive abdominal muscular pressure was applied to the chest wall in those instances. Other studies of non-disabled individuals have indicated that it is common for speech utterances to extend below REL (Hoit & Hixon, 1987; Hoit et al., 1989; Hoit et al., 1996; Winkworth & Davis, 1997), especially when individuals produce a higher number of syllables per breath group (Hodge & Rochet, 1989). With respect to individuals with SCI, it has been reported that they tend to terminate expiratory phases of speech breathing at higher volumes than individuals without SCI (Hoit et al., 1990). This tendency was reflected in the present investigation. In the group of individuals with SCI, there is only one instance where a participant spoke below REL (EJ; Figure 4-16). Examination of Figure 4-16 reveals that this individual with SCI spoke below REL only when dictating the California Passage with continuous-speech dictation software. This particular individual had an incomplete lesion, which may account for his ability to speak below REL using spared thoracic and/or abdominal musculature. However, caution in interpreting this result for EJ is warranted because the possibility exists that an undetected postural shift may be responsible for the placement of this trace. Finally, observation of the kinematic data reveals that lung volume initiation levels were very similar between speaker groups.

With respect to dictation condition, differences were noted with respect to the lung volume level at which speech was initiated and terminated. Tables 4-2 through 4-5 contain information regarding initiation and termination levels across the three dictation conditions for individuals with and without SCI. Table 4-2 contains information regarding lung volume initiation levels for the control subjects. Each speaker's relative volume chart (Figures 4-3 to 4-17) is characterized by 6 speech-breathing traces (2 speech samples for each of 3 dictation conditions). Therefore, for 6 control subjects, there were 36 lung volume initiation levels that could be described. Observation of Table 4-2 reveals that 13/36 times (36%), lung volume initiation levels for individuals in the control group were greater than 1 litre above REL. Lung volume initiation levels between REL and 1 litre above REL occurred 23/36 times (64%). No lung volume initiations were made below REL for this group of subjects. Thus, for the control subjects in this study, higher lung volume initiation levels occurred more frequently in the continuous-speech ASR dictation condition than in either the discrete-word or the control condition.

Table 4-3 contains information regarding lung volume initiation levels for the subjects with SCI. For the 5 subjects with SCI, there were 30 lung volume initiation levels that could be described across dictation conditions (data for MH are missing). Observation of Table 4-3 reveals that 8/30 times (27%), lung volume initiation levels for individuals with SCI were greater than 1 litre above REL. Lung volume initiation levels between REL and 1 litre above REL occurred 22/30 times (73%). No lung volume initiations were made below REL for this group of subjects. Thus, as was the case for the control subjects, speakers with SCI used higher lung volume initiation levels more frequently in the continuous-speech ASR dictation condition than in the other two conditions.

Observation of Table 4-4 for control subjects reveals that a lung volume termination level greater than 1 litre above REL was recorded only once in the 36 breath groups observed (3%). Lung volume termination levels between REL and 1 litre above REL occurred 16/36 times (44%), and termination levels below REL occurred 19/36 times (53%). Thus, for the control subjects in this study, lung volume termination levels below REL occurred most frequently in the control dictation condition. In the discrete-word condition, the majority of terminations occurred between REL and 1 litre above REL. In the continuous-speech ASR condition, there was a more even spread of terminations between lung volume levels that were either below REL or between REL and 1 litre above.

Observation of Table 4-5 for subjects with SCI reveals that lung volume termination levels between REL and 1 litre above REL occurred 29/30 times (97%), and only a single termination level below REL was recorded. Lung volume termination levels occurred most frequently between REL and 1 litre above REL for all dictation conditions. As was the case for the control subjects, the majority of terminations in the discrete-word condition occurred between REL and 1 litre above REL while in the continuous-speech ASR condition, there was a more even spread of terminations between lung volume levels that were either below REL or between REL and 1 litre above.

Lung Volume Initiation Level (Control Subjects)	no ASR	continuous-speech ASR	discrete-word ASR	Row Totals
Greater than 1 litre above REL	2 (17%)	8 (67%)	3 (25%)	13 (36%)
Between REL and 1 litre above REL	10 (83%)	4 (33%)	9 (75%)	23 (64%)
Below REL	0	0	0	0

Table 4- 2. Lung-volume initiation levels for the control subjects across dictation conditions. A total of 36 initiation levels were recorded for the 6 control subjects. The number in each cell of the table represents the frequency with which initiation levels occurred relative to REL.

Lung Volume Initiation Level (SCI)	no ASR	continuous-speech ASR	discrete-word ASR	Row Totals
Greater than 1 litre above REL	2 (20%)	4 (40%)	2 (20%)	8 (27%)
Between REL and 1 litre above REL	8 (80%)	6 (60%)	8 (80%)	22 (73%)
Below REL	0	0	0	0

Table 4- 3. Lung-volume initiation levels for the subjects with SCI across dictation conditions. A total of 30 initiation levels were recorded for the 5 subjects with SCI. The number in each cell of the table represents the frequency with which initiation levels occurred relative to REL.

Lung Volume Termination Level (Control Subjects)	no ASR	continuous-speech ASR	discrete-word ASR	Row Totals
Greater than 1 litre above REL	0	1 (8%)	0	1 (36%)
Between REL and 1 litre above REL	2 (17%)	5 (42%)	9 (75%)	16 (44%)
Below REL	10 (83%)	6 (50%)	3 (25%)	19 (53%)

Table 4- 4. Lung-volume termination levels for the control subjects across dictation conditions. A total of 36 termination levels were recorded for the 6 control subjects. The number in each cell of the table represents the frequency with which termination levels occurred relative to REL.

Lung Volume Termination Level (SCI)	no ASR	continuous-speech ASR	discrete-word ASR	Row Totals
Greater than 1 litre above REL	0	0	0	0
Between REL and 1 litre above REL	10 (100%)	9 (90%)	10 (100%)	29 (97%)
Below REL	0	1 (10%)	0	1 (3%)

Table 4- 5. Lung-volume termination levels for the subjects with SCI across dictation conditions. A total of 30 termination levels were recorded for the 5 subjects with SCI. The number in each cell of the table represents the frequency with which termination levels occurred relative to REL.

With respect to lung volume initiation levels, it appeared that levels of lung volume initiation for speech were higher for dictation with the ASR systems, especially with continuous-speech ASR, than for the control condition. It could be theorized that lung volume initiations were higher in the ASR conditions if speech during dictation was produced with a loud voice (Hixon, Goldman & Mead, 1973). It is possible that vocal loudness may increase if a speaker reacts emotionally to their speech not being recognized by the ASR system. Emotional states, especially those of anger and frustration, tend to increase the loudness of the voice (Leinonen, Hiltunen, Linnankoski & Laakso, 1997). Thus, the anticipation for greater respiratory demands for the production of a louder voice may prompt speakers to initiate utterances at a higher lung volume level than they would for speech that is produced at a natural loudness level (Stathopoulos & Sapienza, 1993).

Differences in lung-volume termination levels were observed across dictation conditions. Most of the breath groups within this study that were produced in the discrete-word condition consisted of less than six syllables and terminated above REL in the control group. In another study of speech-breathing behavior in non-disabled individuals, Hodge and Rochet (1989) found that breath groups during spontaneous speech that contained less than six syllables terminated at or above REL. Within the current investigation, speech produced by the control group in the continuous-speech ASR condition of this study was characterized by approximately 4-8 syllables per breath group, with an equal number of breath groups terminated above REL as below it. Hodge and Rochet also found that for breath groups between 6 and 10 syllables, an equal number terminated below REL as above it. Finally, the majority of breath groups that were produced in the control condition by the non-disabled individuals within this study consisted of a larger number of syllables, and the majority of them terminated below REL. Likewise, Hodge and Rochet found that for breath groups with larger numbers of syllables tended to terminate below REL. Thus, it appears that the regulating influence of ASR on syllable output appears to result in changes in termination levels that would be expected for non-disabled individuals. In individuals with SCI, however, the termination levels were influenced to a greater degree by their impairment as they did not speak below REL, except in one rare circumstance.

Differences in initiation and termination levels relative to REL between speaking tasks of reading and conversation also have been reported in the literature. Hodge and Rochet (1989) reported slightly higher mean end-inspiratory and end-expiratory levels relative to REL during reading than during conversational speech. Within the current study, initiation and termination of lung volume levels were variable across speaking tasks, with no specific patterns in initiation or termination levels relative to REL emerging for dictation of either the letter or California Passage. Therefore, it appears that the speaking task did not interact with the point at which speech was initiated or terminated relative to REL.

To summarize, the results for lung volume initiation and termination levels revealed some differences between groups and across dictation conditions, but essentially no differences between speech samples. The results for lung volume initiation and termination levels confirm that lung volume termination levels are affected by the neuromuscular integrity of the respiratory system and that the regulatory effects of ASR may prompt speakers to alter initiation levels, especially when dictating with continuous-speech ASR.

Lung Volume Excursions

The kinematic data in Figures 4-3 through 4-17 were visually examined to verify that the lung volume excursions represented by each speech-breathing trace reflected the data reported in Chapter 3 for average volume expenditure for speech. Comparisons between groups, across dictation conditions, and between speech samples were considered and will be discussed next.

Although not formally measured using the relative volume charts, the volume expenditures exemplified in the speech-breathing traces generally appeared to be smaller for the group of individuals with SCI than for those without SCI. This reflects the observed trends in the average absolute volume expenditure per expiratory segments of speech that were observed between groups and reported in Chapter 3. In addition, observation of the relative volume charts also reveals differences in volumes exchanged across dictation conditions regardless of speaker group or speech sample. For most of the participants within this investigation, differences in the volume exchanged across dictation conditions are most obvious between dictation in the control condition and dictation with discrete-word ASR. As can be seen in the relative volume diagrams (Figures 4-3 through 4-17), lung volume excursions are generally the greatest when subjects were dictating in the control condition, and the least when they dictated with discrete-word ASR. These kinematic data are consistent with the results for average absolute lung volume exchange reported in Chapter 3. Exceptions to this pattern were subjects CM (Figure 4-7), JA (Figure 4-10), and GC (Figures 4-12 through 4-14). For these three subjects with SCI, there was very little difference in the relative volume excursions of the speech-breathing traces amongst dictation conditions. Finally, as was reported in Chapter 3 of this study, there were no apparent differences in lung volume excursions between the reading and spontaneous speaking tasks. These findings agree with other findings reported in the literature (Hodge & Rochet, 1989), and confirm the results reported in Chapter 3 of no difference in lung volume excursions between the two speech tasks.

In summary, the results for lung volume excursions revealed some differences between groups and across dictation conditions, but essentially no differences between speech samples. The kinematic data collected for individuals with SCI in this study reflect speech-breathing behaviors that would be expected when dealing with respiratory systems characterized by shallow inspirations and volume compression limitations (Hixon & Putnam, 1983; Hoit et al., 1990). With regard to dictation conditions, it appears that the type of speech input that each ASR system requires for successful recognition will have a bearing on the excursions made in anticipation of speech production. Thus, the results for lung volume excursions confirm that the neuromuscular integrity of the respiratory system affects the extent of inspiratory and expiratory events, and that the regulatory effects of ASR may prompt speakers to alter the amount of air that they make available for speech based upon the linguistic demands of the ensuing utterance.

Background Configuration of the Chest Wall

Figures 4-3 through 4-17 represent the respiratory kinematic data collected for the individuals in this study and traced from the digital storage oscilloscope in the form of relative volume charts. Due to the complexity of their traces and the overlap of tracings that would have occurred on one figure, tracings for two subjects with SCI, GC and PM, are shown in three separate figures. All subjects are represented except for one subject with SCI (MH) whose data could not be used due to baseline instability.

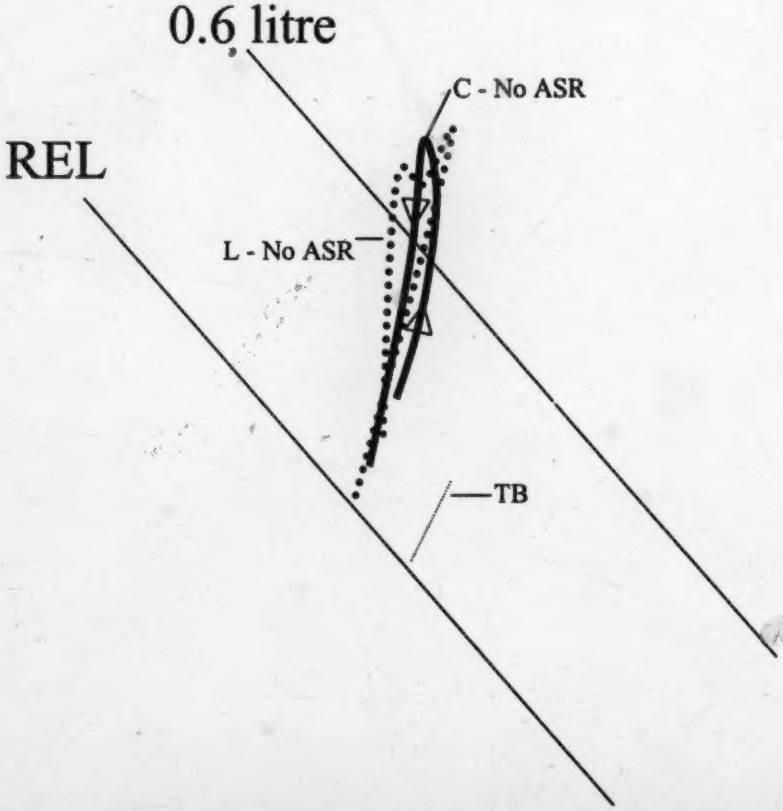

Figure 4- 3. Kinematic tracings for a subject with SCI (PM) during dictation of a letter (L) and the California Passage (C) with no ASR. Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). This individual exhibits a looped pathway for speech breathing that follows a counterclockwise course. The right side of the loop denotes inspiration (△), and the left side denotes expiration (▽). Speech breathing traces lie to the left of the tidal breathing line. The subject tended to terminate speech near REL. This participant's vital capacity was 2.3 litres.

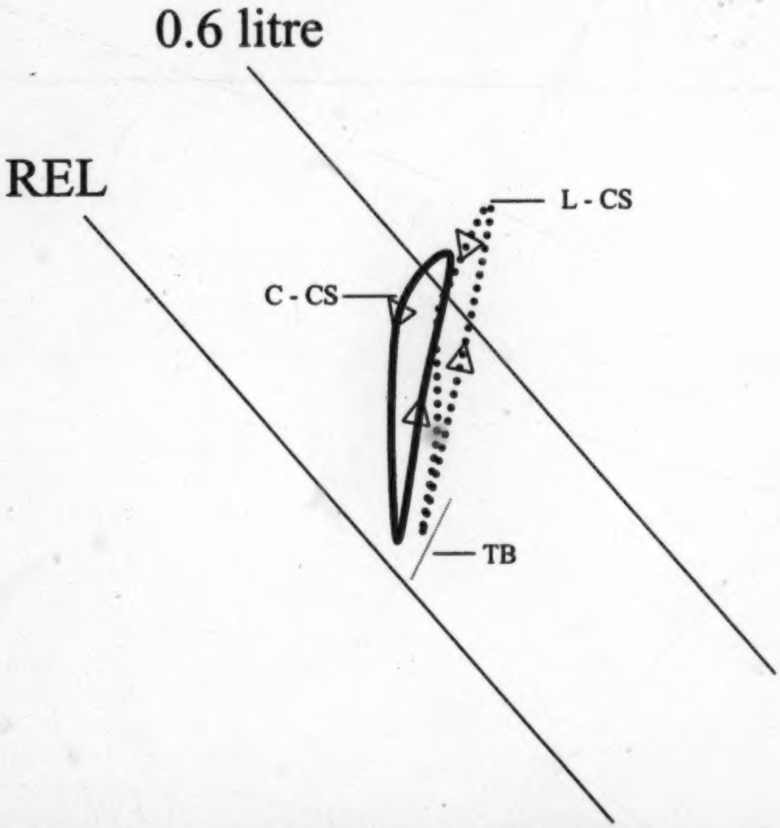

Figure 4- 4. Kinematic tracings for a subject with SCI (PM) during dictation of a letter (L) and the California Passage (C) with continuous-speech ASR (CS). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). This individual exhibits a looped pathway for speech breathing that follows a counterclockwise course. The right side of the loop denotes inspiration (△), and the left side denotes expiration (▽). Speech breathing traces lie to the left of the tidal breathing line. The subject tended to terminate speech near REL. This participant's vital capacity was 2.3 litres.

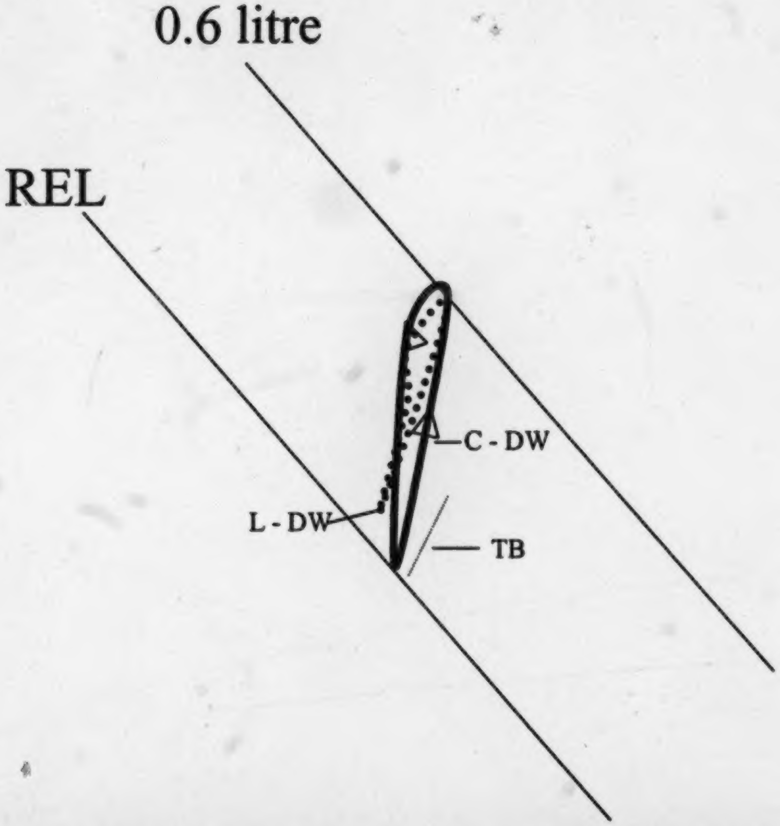

Figure 4- 5. Kinematic tracings for a subject with SCI (PM) during dictation of a letter (L) and the California Passage (C) with discrete-word ASR (DW). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). This individual exhibits a looped pathway for speech breathing that follows a counterclockwise course. The right side of the loop denotes inspiration (△), and the left side denotes expiration (▽). Speech breathing traces lie to the left of the tidal breathing line. The subject tended to terminate speech near REL. This participant's vital capacity was 2.3 litres.

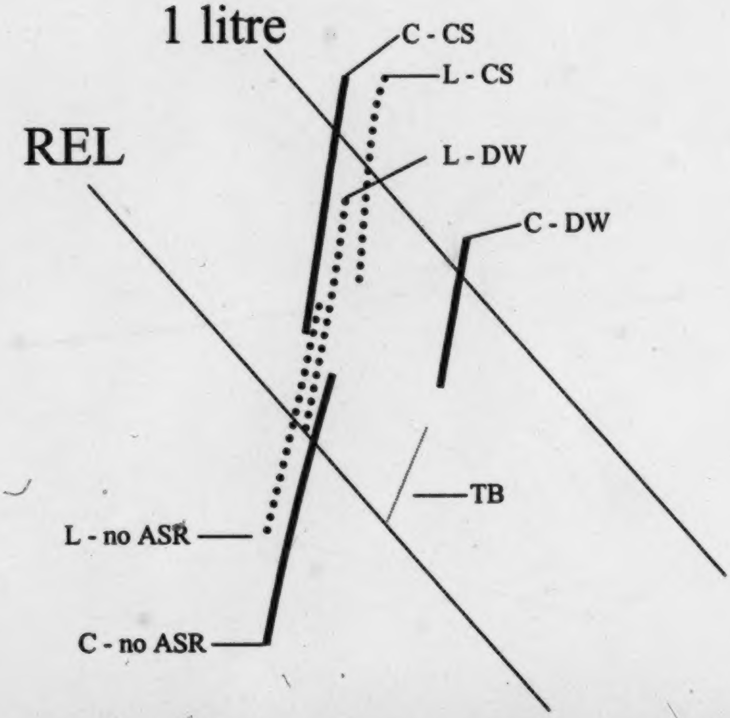

Figure 4- 6. Kinematic tracings for a control subject (KT) during dictation of a letter (L) and the California Passage (C) in all three dictation conditions (no ASR; continuous-speech [CS]; discrete-word [DW]). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). This individual exhibits single, relatively straight pathways for both inspiration and expiration. Most of the speech breathing traces lie to the left of the tidal breathing line. The C-DW line appears to lie on the same path as the tidal breathing line. This individual initiated and terminated speech at a higher lung volume when dictating with both of the ASR systems than when dictating without them. In addition, the subject tended to speak near or below REL when dictating with no ASR. This participant's vital capacity was 3.7 litres.

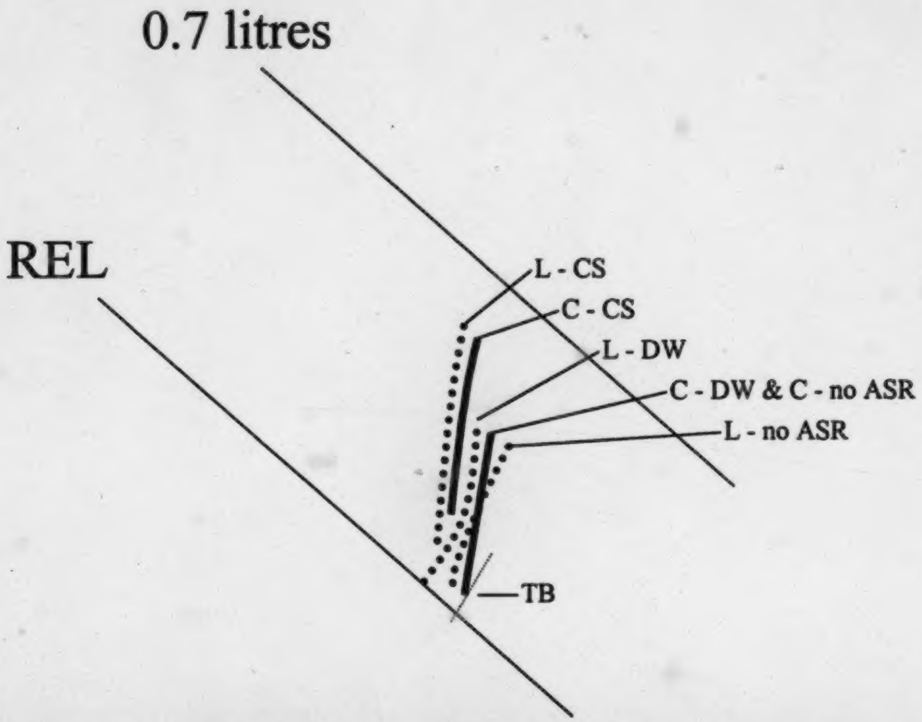

Figure 4- 7. Kinematic tracings for a subject with SCI (CM) during dictation of a letter (L) and the California Passage (C) in all three dictation conditions (no ASR; continuous-speech [CS]; discrete-word [DW]). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). This individual exhibits single, relatively straight pathways for both inspiration and expiration. All of the speech breathing traces lie to the left of the tidal breathing line. This individual terminated speech above or near REL in all dictation conditions. This participant's vital capacity was 2.3 litres.

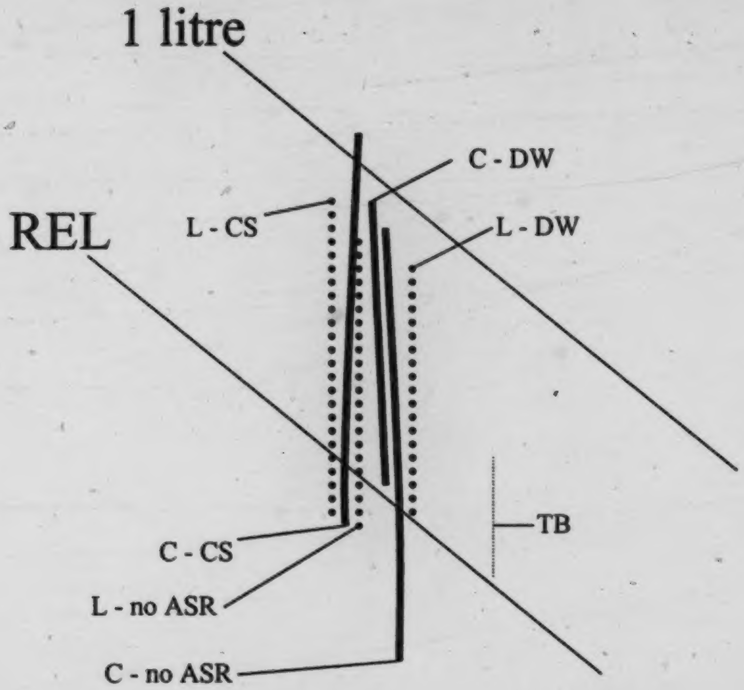

Figure 4- 8. Kinematic tracings for a control subject (FM) during dictation of a letter (L) and the California Passage (C) in all three dictation conditions (no ASR; continuous-speech [CS]; discrete-word [DW]). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). This individual exhibits single, relatively straight pathways for both inspiration and expiration. All of the speech breathing traces lie to the left of the tidal breathing line. This individual terminated speech above or near to REL only when dictating in the discrete-word condition. Speech was terminated below REL for all other conditions. This participant's vital capacity was 3.5 litres.

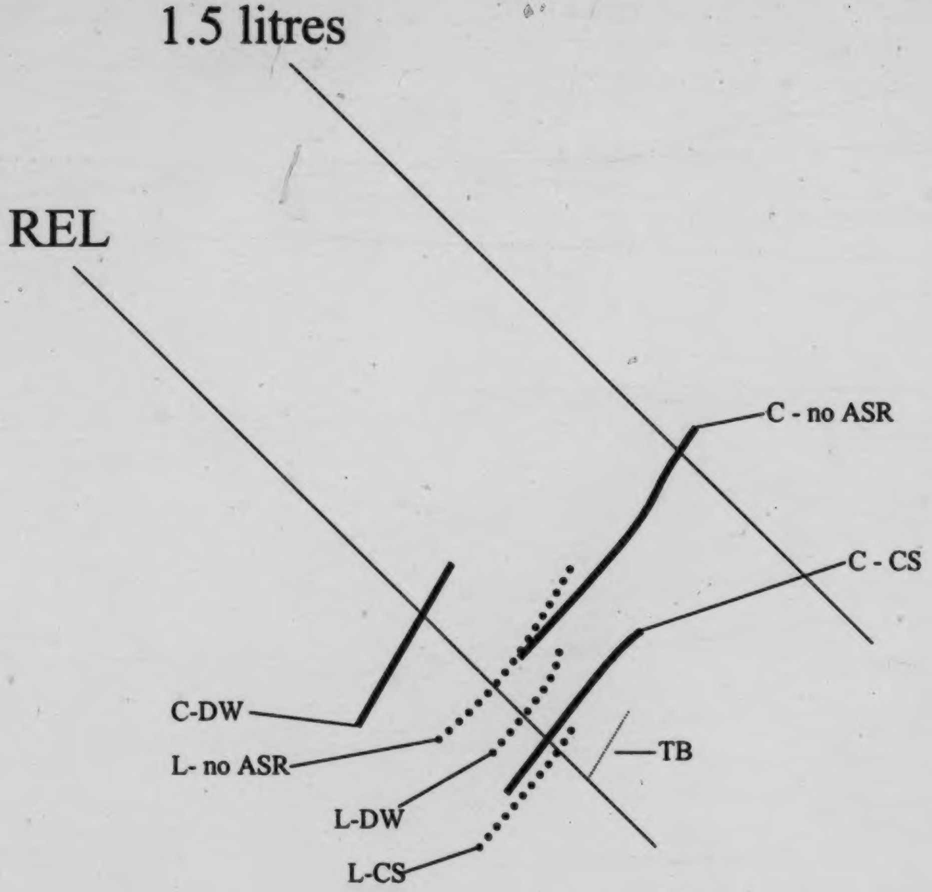

Figure 4- 9. Kinematic tracings for control subject (TA) during dictation of a letter (L) and the California Passage (C) in all three dictation conditions (no ASR; continuous-speech [CS]; discrete-word [DW]). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). This individual exhibits single, relatively straight pathways for both inspiration and expiration. All of the speech breathing traces lie to the left of the tidal breathing line. This individual terminated speech above REL only when dictating the California Passage in the no ASR condition. Speech was terminated below REL for all other conditions. This participant's vital capacity was 4.3 litres.

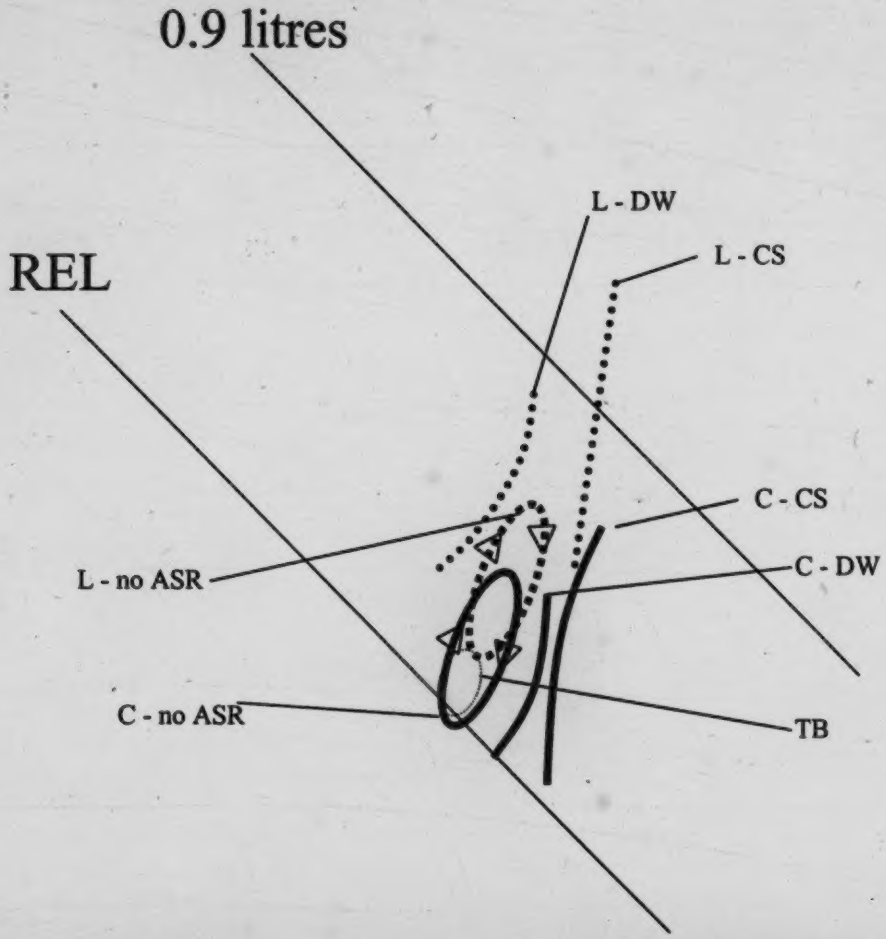

Figure 4- 10. Kinematic tracings for a subject with SCI (JA) during dictation of a letter (L) and the California Passage (C) in all three dictation conditions (no ASR; continuous-speech [CS]; discrete-word [DW]). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted oval line represents tidal breathing (TB). This individual exhibits a looped pathway that follows a clockwise course for tidal breathing and for speech breathing in the no ASR dictation condition. The left side of the loop denotes inspiration (△), and the right side denotes expiration (▽). For the rest of the dictation conditions, this individual exhibits single, relatively straight pathways for both inspiration and expiration. The speech breathing traces representative of dictation with no ASR lie on the same path as that of the tidal breathing lines. The trace for dictation of a letter in the discrete-word condition lies to the left of the tidal breathing line. The rest of the traces lie to the right of the tidal breathing line. This individual terminated speech above REL in all dictation conditions. This participant's vital capacity was 2.3 litres.

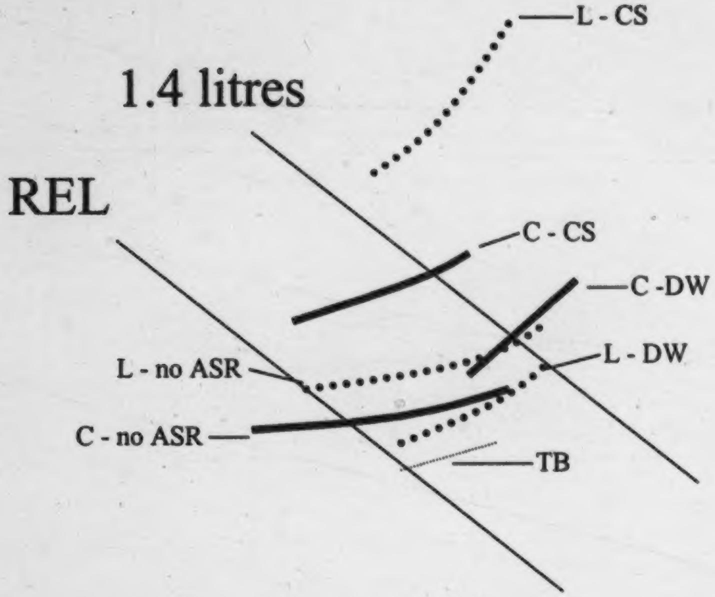

Figure 4- 11. Kinematic tracings for a control subject (BS) during dictation of a letter (L) and the California Passage (C) in all three dictation conditions (no ASR; continuous-speech [CS]; discrete-word [DW]). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). This individual exhibits single, relatively straight pathways for both inspiration and expiration. All of the speech breathing traces lie to the left of the tidal breathing line. This individual terminated speech above REL in all conditions, except for dictation of the California Passage with no ASR. This participant's vital capacity was 3.3 litres.

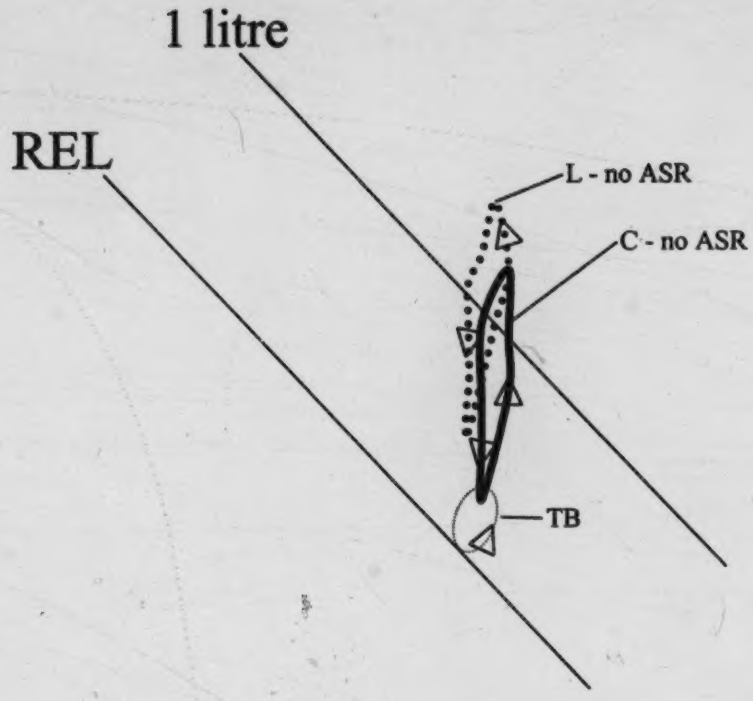

Figure 4- 12. Kinematic tracings for a subject with SCI (GC) during dictation of a letter (L) and the California Passage (C) with no ASR. Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted oval line represents tidal breathing (TB). This individual exhibits looped pathways for tidal breathing and speech breathing that follow a counterclockwise course. The right side of the loop denotes inspiration (△), and the left side denotes expiration (▽). The speech breathing traces lie slightly to the left of the intersection of the tidal breathing loop with REL. This subject terminated speech above or near REL across all dictation conditions. This participant's vital capacity was 1.7 litres.

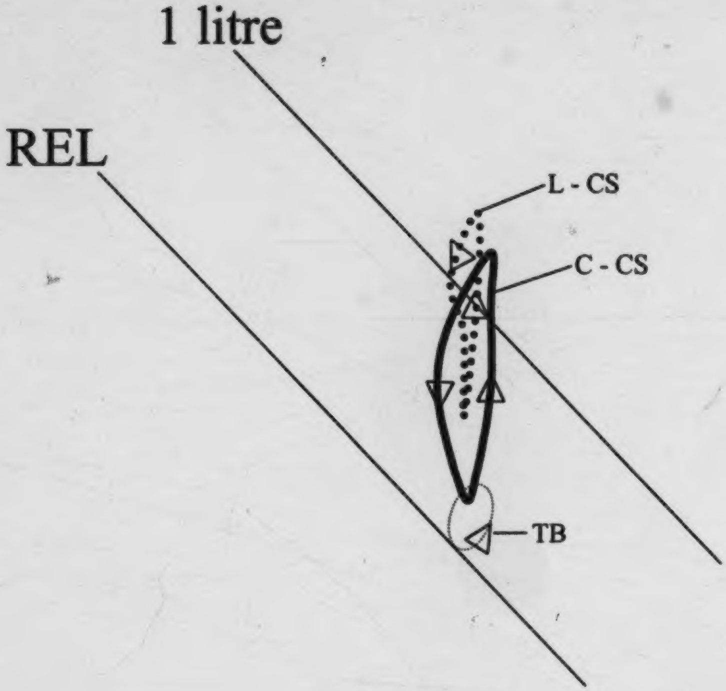

Figure 4- 13. Kinematic tracings for a subject with SCI (GC) during dictation of a letter (L) and the California Passage (C) with continuous-speech ASR (CS). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted oval line represents tidal breathing (TB). This individual exhibits looped pathways for tidal breathing and speech breathing that follow a counterclockwise course. The right side of the loop denotes inspiration (△), and the left side denotes expiration (▽). The speech breathing traces lie slightly to the left of the intersection of the tidal breathing loop with REL. This subject terminated speech above or near REL across all dictation conditions. This participant's vital capacity was 1.7 litres.

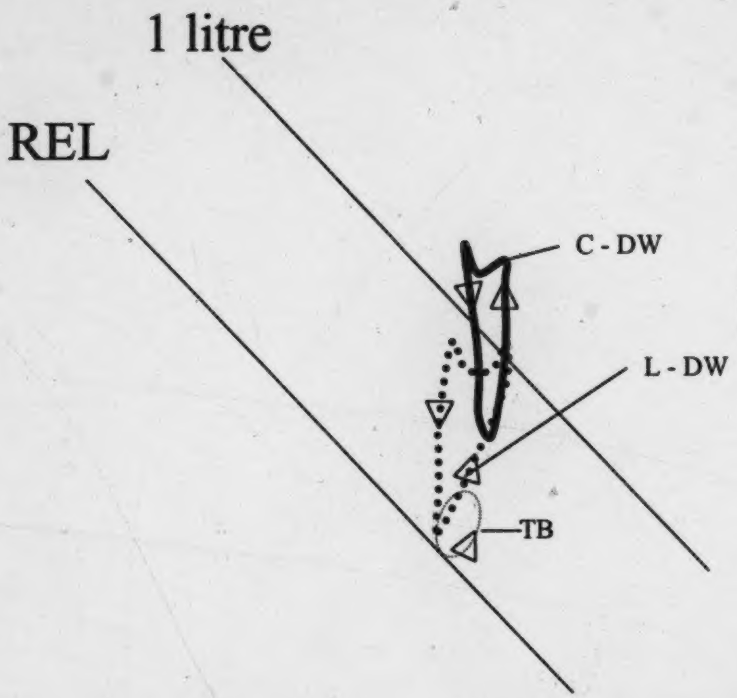

Figure 4- 14. Kinematic tracings for a subject with SCI (GC) during dictation of a letter (L) and the California Passage (C) with discrete-word ASR (DW). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted oval line represents tidal breathing (TB). This individual exhibits looped pathways for tidal breathing and speech breathing that follow a counterclockwise course. The right side of the loop denotes inspiration (△), and the left side denotes expiration (▽). This subject terminated speech above or near REL across all dictation conditions. This participant's vital capacity was 1.7 litres.

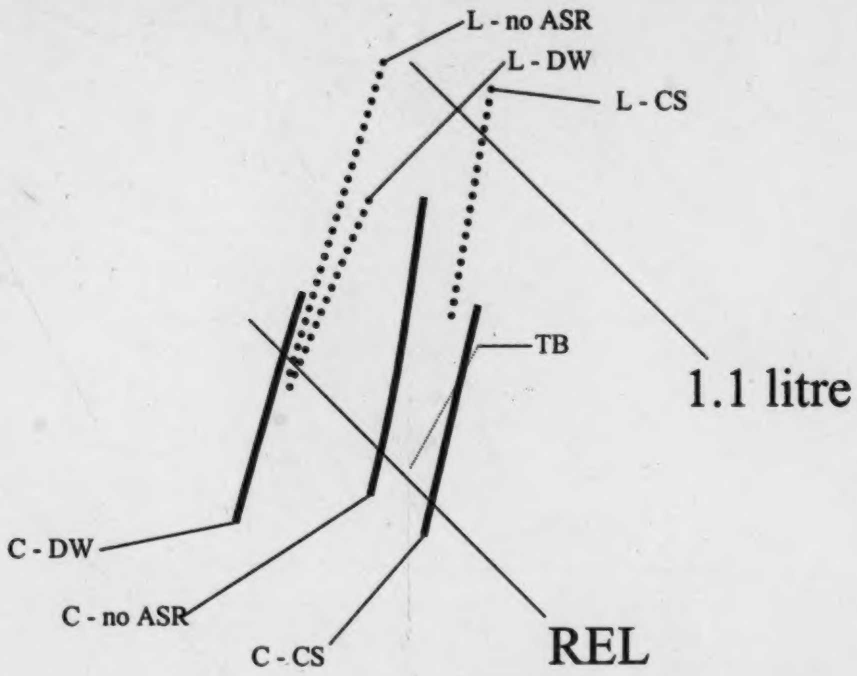

Figure 4- 15. Kinematic tracings for a control subject (RT) during dictation of a letter (L) and the California Passage (C) in all three dictation conditions (no ASR; continuous-speech [CS]; discrete-word [DW]). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). This individual exhibits single, relatively straight pathways for both inspiration and expiration. All but one of the speech breathing traces lie to the left of the tidal breathing line. The speech-breathing trace representing dictation of the California Passage in the continuous-speech dictation condition begins on the left of the tidal breathing line and terminates to the right of it. This individual terminated speech below REL when dictating the California Passage in all three dictation conditions. Speech was terminated at or above REL when spontaneously dictating a letter, regardless of dictation condition. This participant's vital capacity was 4.4 litres.

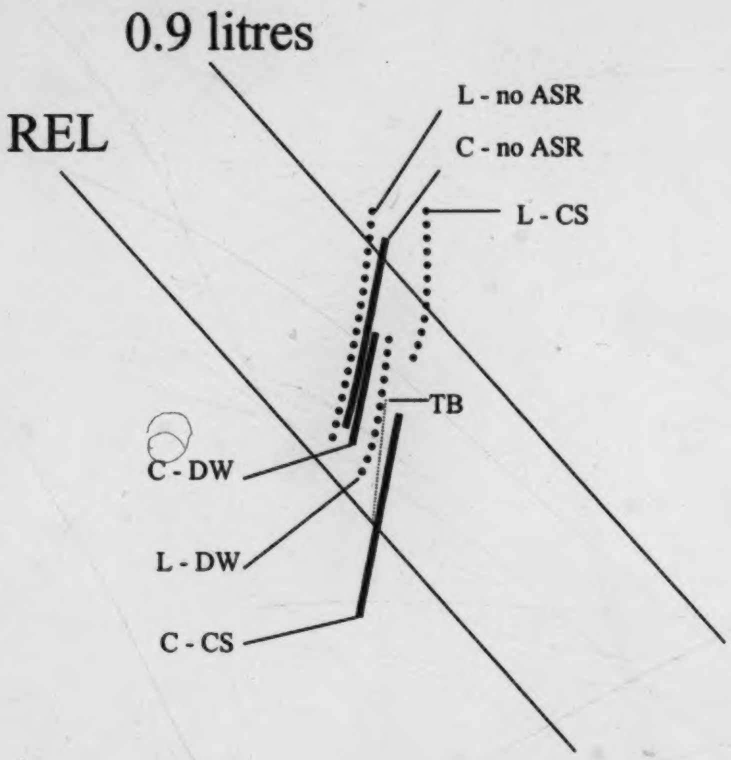

Figure 4- 16. Kinematic tracings for a subject with SCI (EJ) during dictation of a letter (L) and the California Passage (C) in all three dictation conditions (no ASR; continuous-speech [CS]; discrete-word [DW]). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). All but two of the speech breathing traces lie to the left of the tidal breathing line. The traces representing dictation with continuous-speech ASR lie slightly to the right of the tidal breathing line. This individual terminated speech below REL only when dictating the California Passage in the continuous-speech ASR condition. This participant's vital capacity was 3.2 litres.

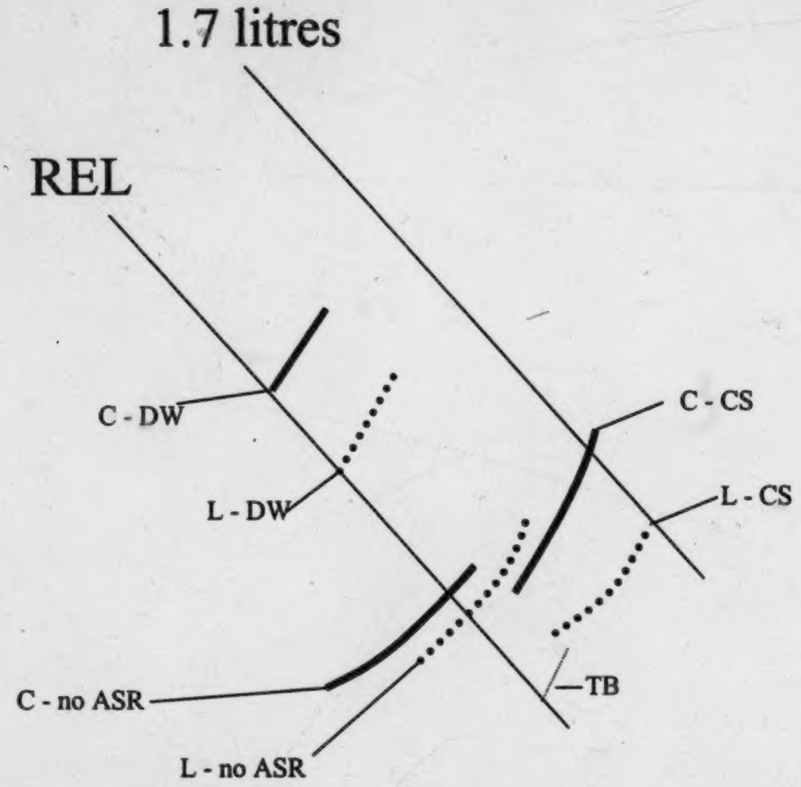

Figure 4- 17. Kinematic tracings for a control subject (GG) during dictation of a letter (L) and the California Passage (C) in all three dictation conditions (no ASR; continuous-speech [CS]; discrete-word [DW]). Each pattern represents a composite of approximately 45 seconds of speech. The thin dotted line represents tidal breathing (TB). All of the speech breathing traces lie to the left of the tidal breathing line. This individual only spoke below REL when dictating with no ASR. All other speech breathing traces were terminated either above or at REL. This participant's vital capacity was 5.8 litres.

The following discussion of the background configuration of the chest wall will include description of the speech-breathing trace departures from the 'intersection of the tidal-breathing line with REL' and the position of the tidal breathing line, which will be presumed to represent a configuration close to relaxation. Description of the characteristics of background configuration of the chest wall is based on information in the relative volume charts in Figures 4-3 through 4-17. Background configuration will be described for each experimental condition in the following order: between speaker groups, across dictation conditions, and between speech samples.

Speaker Group

With respect to departures from the presumed relaxation configuration, similarities and differences in background configuration of the chest wall existed between the two groups of individuals within this study. One similarity between groups was that the majority of the speech-breathing traces were situated either on or to the left of the tidal breathing lines. Three of the 5 individuals with SCI and all of the control subjects exhibited chest wall configurations that were consistent with normal torso configuration for speech. Exceptions to this were exemplified by two individuals with SCI (JA, Figure 4-10; EJ, Figure 4-16). Observation of Figures 4-10 and 4-16 reveal that 4 out of the 12 speech breathing traces were situated to the right of the tidal breathing lines for these two participants. As indicated earlier in the chapter, speech breathing traces that lie to the left of the tidal breathing line are consistent with normal torso configuration for speech in the upright position where the rib cage is larger and the abdomen is smaller than their relaxed configurations at the same lung volume (Hixon, Goldman & Mead, 1973). That the subjects with SCI exhibited some speech-breathing traces that fell slightly to the left of the presumed relaxation point suggests that spared rib cage muscles, such as the pectoralis major and latissimus dorsi, and neck muscles, such as the sternocleidomastoid, scalenes and subclavius, may have been active for maintenance of this posture (Hixon & Putnam, 1983; Hoit et al., 1990). Electromyographic studies have revealed that the pectoralis major, latissimus dorsi and sternocleidomastoid are active respiratory muscles in individuals with quadriplegia (Fujiwara, Hara & Chino, 1999). Thus, the data placements for the subjects with SCI suggest that muscles such as these were activated to provide a pull on the rib cage, thereby increasing its circumference and creating a leftward departure of the speech-breathing traces from the presumed relaxation characteristic during speech activities (Hixon, Putnam & Sharp, 1983).

Although the majority of the speech-breathing traces for individuals in each group fell on or to the left of the presumed relaxation point, there were subtle differences in the degree to which they did so. For example, most of the utterances for individuals with SCI clustered around the presumed relaxation line, whereas the utterances for individuals in the control group were more likely to depart farther to the left of the presumed relaxation point. In the group of individuals with SCI, two exceptions to the pattern of clustering around the presumed relaxation point were seen in the relative volume charts for subjects PM (Figure 4-3) and CM (Figure 4-7). Subject PM's speech-breathing traces departed to the left from the presumed relaxation point, especially during dictation in the control condition. Subject CM's speech-breathing traces departed to the left of the presumed relaxation line, especially during dictation with continuous-speech ASR. That both of these subjects' speech-breathing traces departed to the left of the presumed relaxation point, as do those of most non-disabled subjects, suggests that these two individuals with SCI may have been able to recruit accessory rib cage muscles to a greater degree than the other individuals with SCI. Because the speech-breathing traces for the other individuals with SCI clustered around the relaxation characteristic, it appears that those speakers tended to assume a chest wall shape that required little active muscular pressure. It is possible that the individuals who's speech-breathing traces clustered around the relaxation line assumed this configuration because they did not have the same muscular ability to distort their chest wall from the relaxation characteristic as the other individuals with SCI. It is also possible that they were reserving the muscular energy that would be required to set the chest wall in this configuration for other respiratory efforts such as inhalation and possible checking action during expiration.

Dictation Condition

For most subjects, speech-breathing data clustered in the same area within the relative volume chart regardless of dictation condition. Occasionally for some subjects, one trace would fall to the left or to the right of the cluster of other speech-breathing traces in the relative volume chart. An example of this is illustrated in the relative motion chart for control subject KT (Figure 4-6) for whom the speech-breathing trace for dictation of the California passage with discrete-word ASR fell to the right of the cluster of other speech-breathing traces, indicating that this subject assumed a less prominent inverted-pear shape during dictation with discrete-word ASR. For control subject TA (Figure 4-9), an opposite pattern was seen. For this subject, the speech-breathing trace for dictation of the California Passage with discrete-word ASR fell to the left of the other speech-breathing traces, indicating a further departure from the presumed relaxation point for dictation with discrete-word ASR than for continuous-speech ASR or natural dictation. Some caution in the interpretation of the difference in positioning of the discrete-word trace from the other speech-breathing traces observed for KT and TA should be exercised due to the possibility that such placement could also have been due to postural shifts in each subject that went undetected. This may be especially apparent for subject KT whose 'C-No ASR' breathing trace extends unusually far below REL, suggesting that undetected postural shifts were likely.

In only two cases did a pair of speech-breathing traces representing both speech tasks for one dictation condition cluster together and away from the other traces. This can be seen in the relative volume charts for control subjects BS (Figure 4-11) and GG (Figure 4-17). For subject BS, observation of Figure 4-11 reveals that the speech-breathing traces for dictation with continuous-speech ASR occurred to the left of the other speech-breathing traces, which clustered around the tidal-breathing trace. This pattern indicates that this subject assumed a more prominent inverted-pear shape during dictation with continuous-speech ASR than for dictation with discrete-word ASR or in the control condition. On the other hand, Figure 4-17 reveals that the speech-breathing traces of subject GG for dictation with discrete-word ASR occurred to the left of the other speech-breathing traces, which clustered around the tidal-breathing trace. This pattern indicates that this subject assumed a more prominent inverted-pear shape during dictation with discrete-word ASR than for dictation with continuous-speech ASR or in the control condition.

Thus, for some individuals, it appears that dictation with ASR has the potential to influence the configuration of the chest wall in a manner that differs from that associated with dictation without ASR. However, there are no consistent trends in departures from the control speaking condition based on type of ASR and therefore, no definite conclusions can be drawn about the influence of different types of ASR on chest wall configuration.

Speech Sample

Departures from the presumed chest wall relaxation configuration represented by tidal-breathing traces did not follow any specific pattern during the reading task versus the spontaneous-speaking task. For any individual, sometimes the traces for dictation of a letter fell further to the left of the presumed relaxation line than traces for dictation of the California Passage in the same dictation condition, and other times the converse was true. Thus, it appears that the speaking task was not an important variable in predicting the direction or extent of departure of the speech-breathing traces from the presumed relaxation point.

In summary, differences in background configuration of the chest wall were observed between speaker groups, but not clearly among dictation conditions nor between speech samples. Thus, it appears that the integrity of the respiratory system affects background configuration during speech, leading to differences in background chest wall shape between individuals with and without SCI. The influence of ASR on background configuration is less clear as there were no consistent trends revealed across dictation conditions. Finally, the linguistic differences between the two speaking tasks did not provoke differences in speech-breathing behavior that was reflected in the kinematic data within the current investigation.

Kinematic Data Concluding Remarks

The kinematic data collected in this investigation reveal that differences were most obvious between speaker groups and dictation conditions, but not between speech samples. With respect to speaker group, differences were observed with respect to initiation and termination of speech relative to REL, lung volume excursions, and background configuration of the chest wall. However, there were essentially no differences in the %RC contribution between individuals with and without SCI suggesting that individuals with SCI recruited accessory rib cage and neck muscles to affect volume changes. That such muscles could be recruited was confirmed through the observation that these individuals produced speech-breathing traces that fell to the left of the presumed relaxation characteristic, a feat that requires the muscles of the rib cage or neck to be called into action to expand the rib cage.

With respect to dictation conditions, differences were also observed across conditions in initiation and termination of speech relative to REL, and in lung volume excursions. Although some differences were observed in the background configuration of the chest wall across dictation conditions, they did not present any particular pattern. The results for %RC contribution across dictation conditions suggested that there were some differences; however, the differences were small and fall within the values recorded for measurement error and therefore, interpretation of this finding is guarded.

Regarding speech sample, no clear trends emerged for any of the variables that were described. Thus, it appears that the linguistic differences between the two speaking tasks did not lead to differences in the kinematic data within the current investigation.

The efficiency issues that arose with respect to the speech-breathing data in Chapter 3 can be carried forward to this chapter and expanded upon. With respect to rib cage contribution, it appears that most individuals, regardless of speaker group, used the rib cage to effect lung volume exchange. As stated earlier, the rib cage covers a larger surface of the pulmonary system than the abdomen and therefore, may be a more efficient producer of lung volume change than the abdomen (Hixon, Goldman & Mead, 1973; Hixon, Mead & Goldman, 1976; Hoit, 1995; Hoit et al., 1996). Thus, it appears that the individuals within this study favored an efficient set up for lung volume exchange. With respect to lung volume initiation and termination, it appears that most speakers were working in the midvolume range of their vital capacity, a range that is characterized by more compliance and less stiffness than the extremes of that capacity (Hixon, Goldman & Mead, 1973). Speech produced outside this range will require greater expenditures of muscular energy (Hixon, Goldman & Mead, 1973). Therefore, it appears that the subjects within this study continued to produce speech in a range that maintained efficient use of muscular energy. Considering lung volume excursions, the kinematic data confirm the efficiency issues related to small inspirations that were brought forth in Chapter 3; less speech will be produced on smaller inspirations associated with ASR dictation thereby requiring more respiratory pump cycles to dictate a message. Therefore, inefficiency results from the regulating effect of ASR on lung volume excursions. Finally, observation of changes in background configuration of the chest wall was not revealing across dictation conditions, but provided some insight to differences in respiratory efficiency between groups. Individuals with SCI were less likely to depart from the relaxation characteristic, with one interpretation of this outcome being related to efficiency for that population. It is possible that individuals with SCI were less likely to depart from the presumed relaxation characteristic because they were reserving the muscular energy that it would take to set the respiratory system up in that configuration for other respiratory efforts such as inhalation and possible checking action during expiration.

In conclusion, differences between speaker groups and across dictation conditions were observed in the kinematic data. Group differences that were exhibited contrast the respiratory limitations of individuals with SCI with the healthy respiratory systems of non-disabled individuals. The volume-related kinematic observations confirm the trends for differences between individuals with and without SCI in volumes exchanged that were revealed in Chapter 3. Differences across dictation conditions reveal the regulatory effect of ASR on speech-breathing behaviors and also confirm the volume exchange results from Chapter 3. Finally, there were no striking differences between speech samples for any of the aforementioned variables.

CHAPTER 5 - Speech Production Results

Purpose

The intent of this portion of the study was to describe a number of characteristics of speech production associated with spontaneous speech and reading aloud in three dictation conditions (with discrete-word ASR, with continuous-speech ASR, and with no ASR) and to determine if speech production variables differed between two groups of users (those with and without spinal cord injury).

Variables

The dependent variables for this portion of the study were as follows:

a. Average number of syllables per breath group (#)

b. Frequency of breath groups per paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)

c. Frequency of apneic episodes per paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)

d. Time needed for dictation of a paragraph read aloud (California Passage) (minutes)

e. Number of words used in dictation of a paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)

Hypotheses & Hypothetical Outcomes

Hypotheses and related hypothetical outcomes that were proposed for the speech production variables of interest are reproduced in Table 5-1.

Dependent Variable	**Speaker Group**	**Dictation Condition**
Average **number of syllables** per breath group (#)…	will be lower for individuals with SCI than for individuals without SCI across dictation conditions	will be lowest when dictating with discrete-word ASR and highest in the no-ASR dictation condition for all speakers
Frequency of breath groups per paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)…	will be higher for individuals with SCI than for those without SCI across dictation conditions	will be highest when dictating with discrete-word ASR and lowest in the no-ASR dictation condition for all speakers
Frequency of **apneic or breath-holding episodes** per paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)…	will be higher for individuals with SCI than for those without SCI across dictation conditions	will be highest during dictation with discrete-word ASR and lowest in the no-ASR dictation condition for all speakers
Time needed for dictation of a paragraph read aloud (California Passage) (minutes)…	will be longer for individuals with SCI than for those without SCI across dictation conditions	will be longest for discrete-word ASR and shortest in the no-ASR dictation condition for all speakers
Number of words used in dictation of a paragraph read aloud (California Passage) and time-limited passage of spontaneous speech (Letter) (#)…	will not differ between individuals with SCI and those without SCI across dictation conditions	will be highest for dictation of the California Passage in the discrete-word ASR condition and lowest in the no-ASR dictation condition for all speakers across dictation conditions; the opposite will be true for the letter

Table 5- 1. Hypothetical outcomes for speech production variables between groups and across dictation conditions.

Method

Data Processing and Analysis

The multichannel data obtained from speakers during the dictation conditions and stored on FM tape served as the source of the speech-production variables. As was the case for analysis of the speech-breathing variables in Chapter 3, the rib cage, abdomen, and speech signals from the FM record were low-pass filtered and digitized at 11kHz and a 12-bit quantization rate for analysis in CSpeech. The three waveform traces (i.e., the microphone signal, rib cage magnetometer signal, and abdomen magnetometer signal) were displayed simultaneously in the same window (see Figure 3-1).

Average number of syllables per breath group

The number of syllables per breath group was determined by locating an expiratory segment that supported speech in the magnetometric data displayed in a

CSpeech window. Within that segment, with reference to the microphone signal, the onset of speech was marked with one cursor and the offset of speech was marked with another cursor. The speech between the two cursors was played and the words in the breath group were identified on the typed transcript of the participant's speech. This process was repeated for each subject's reading and spontaneous speech for all dictation conditions. The number of syllables within each breath group was then tallied and recorded. The total number of syllables per breath group was averaged across the total number of breath groups within each of the two speech samples produced in each of the three dictation conditions.

Frequency of breath groups per sample of read and spontaneous speech

The frequency of breath groups per standard passage was determined by tallying the number of breath groups that were marked during collection of the syllables-per-breath-group data on the typed transcripts for the California passage. For spontaneous dictation of the letter, the number of breath groups produced during the first 3 minutes of dictation was tallied and recorded.

Frequency of apneic episodes per sample of read and spontaneous speech

Occurrences of apnea within a breath cycle that contained speech were tallied during dictation of the California Passage and during the first 3 minutes of spontaneous dictation of a letter. An occurrence of apnea was defined as an instance in which there was no change in the expiratory slopes of the rib-cage and abdominal magnetometer signals for a duration greater than 500 msec. Cursor placements at the onset and offset of such moments in the data displayed in a CSpeech window facilitated application of this duration criterion.

Time needed for dictation of the California Passage

The time needed to dictate the California passage in each condition, including corrections made by the subjects while dictating, was measured with a stop watch while the primary investigator reviewed each participant's taped record.

Number of words used in dictation of read and spontaneous speech

The number of words used in dictation was determined by a word-count function applied to a typed transcript of each subject's dictation of the California Passage and first 3 minutes of spontaneous dictation of a letter. For dictation of the California Passage in the no-ASR condition, the number of words was the number of printed words (220) minus any words that an individual may have skipped, or plus any words that may have been added such as a repetition of a word within the passage. For dictation of the California Passage with the ASR systems, the number of words included all words spoken including commands for computer corrections. For dictation of a letter, the number of words included all words spoken by a participant in the first 3 minutes of dictation.

Statistical Analyses

The speech production variables were analyzed via two series of 2x3 (speaker group x dictation condition) repeated-measures ANOVAs with fixed effects on the first factor and repeated-measures on the second factor. Two series of ANOVAs, rather than one 2x3x2 (speaker group x dictation condition x speech sample) ANOVA, were undertaken due to the inherent differences between the speech samples (spontaneous versus scripted). Therefore, one series of ANOVAs investigated data for the variables obtained during reading aloud, and the other investigated the data collected during spontaneous speech.

Within an ANOVA series, each of the dependent variables was analyzed via a separate 2x3 analysis, having 2 levels on the first factor (SCI & non-SCI) and 3 levels on the second factor (no ASR, continuous-speech ASR, & discrete-word ASR). For the analyses related to dictation of the California Passage, alpha levels were adjusted from p=.05 to p=.01 according to Bonferroni's correction for 5 planned comparisons (.05/5). For analyses related to dictation of a letter, alpha levels were adjusted from p=.05 to p=.0125 according to Bonferroni's correction for 4 planned comparisons (.05/4) (Portney & Watkins, 2000). Measurement reliability was analyzed via intra-class correlations (ICC) for approximately 25% of the data collected.

Results

The California Passage

Table 5-2 contains mean scores and standard deviations for the dependent variables obtained for each speaker group across the three dictation conditions for the California Passage.

	GROUP	DICTATION CONDITION	Mean	Std. Deviation
syllables per breath group (#)	SCI	discrete-word	1.80	8.86E-02
		continuous-speech	5.42	1.65
		no ASR	13.22	4.86
	non-SCI	discrete-word	2.25	.28
		continuous-speech	5.89	2.29
		no ASR	21.27	11.64
frequency of breath groups per standard passage (#)	SCI	discrete-word	331.83	49.74
		continuous-speech	126.17	41.88
		no ASR	29.83	12.70
	non-SCI	discrete-word	300.67	59.93
		continuous-speech	110.17	33.55
		no ASR	19.33	8.80
frequency of apnea per standard passage (#)	SCI	discrete-word	22.17	12.98
		continuous-speech	10.67	7.61
		no ASR	.00	.00
	non-SCI	discrete-word	12.00	11.82
		continuous-speech	3.33	2.66
		no ASR	.00	.00
time needed for dictation (minutes)	SCI	discrete-word	14.19	3.13
		continuous-speech	8.62	1.98
		no ASR	1.29	9.93E-02
	non-SCI	discrete-word	16.72	4.53
		continuous-speech	7.56	2.12
		no ASR	1.19	7.87E-02
number of words needed for dictation (#)	SCI	discrete-word	397.17	83.52
		continuous-speech	398.83	45.29
		no ASR	222.33	1.97
	non-SCI	discrete-word	458.17	56.92
		continuous-speech	368.50	28.18
		no ASR	217.00	13.39

Table 5- 2. Means and standard deviations for the speech-production variables analyzed during dictation of the California Passage between groups and across dictation conditions.

Within-subjects results from each repeated-measures ANOVA for the speech production variables obtained during dictation of the California Passage are summarized in Table 5-3. Between-groups results are summarized in Table 5-4. The analyses revealed significant dictation condition main effects for all of the dependent variables. No interaction effects were found. In addition, no significant differences were found between speaker groups for any of the speech production variables during dictation of the California Passage. Because no group differences were found, the results that are illustrated in Figures 5-1, 5-3, 5-5, 5-7 and 5-9 represent main effects for dictation condition averaged across all speakers. For the reader who is interested in group means, please refer to Table 5-2.

Univariate Tests

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared	Observed Power
DICTATION CONDITION	syllables/breath group	Greenhouse-Geisser	1515.580	1.050	1443.448	31.690	.000*	.760	.999
	frequency of breath groups	Sphericity Assumed	532257.167	2	266128.583	233.213	.000*	.959	1.000
	frequency of apnea	Greenhouse-Geisser	1770.056	1.259	1406.102	19.321	.000*	.659	.992
	time	Sphericity Assumed	1213.627	2	606.813	110.487	.000*	.917	1.000
	number of words	Sphericity Assumed	288384.000	2	144192.000	70.402	.000*	.876	1.000
DICTATION CONDITION * GROUP	syllables/breath group	Greenhouse-Geisser	115.148	1.050	109.667	2.408	.150	.194	.297
	frequency of breath groups	Sphericity Assumed	687.389	2	343.694	.301	.743	.029	.091
	frequency of apnea	Greenhouse-Geisser	165.167	1.259	131.206	1.803	.206	.153	.257
	time	Sphericity Assumed	20.789	2	10.394	1.893	.177	.159	.346
	number of words	Sphericity Assumed	13366.889	2	6683.444	3.283	.059	.246	.554

Table 5- 3. Summary of ANOVA results for within-subjects results for each speech production variable during dictation of the California Passage. *Indicates a statistically significant result.

Tests of Between-Subjects Effects

Transformed Variable: Average

Source	Measure	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared	Observed Power
GROUP	syllables/breath group	80.690	1	80.690	2.258	.164	.184	.275
	frequency of breath groups	3325.444	1	3325.444	1.440	.258	.126	.193
	frequency of apnea	306.250	1	306.250	3.226	.103	.244	.369
	time	1.863	1	1.863	.223	.647	.022	.071
	number of words	641.778	1	641.778	.254	.625	.025	.074

Table 5- 4. Summary of ANOVA results for between-groups results for each speech production variable during dictation of the California Passage.

Average Number of Syllables per Breath Group

The analysis revealed significant differences in number of syllables spoken per breath group amongst dictation conditions. Bonferroni post-hoc analyses revealed significant differences between all dictation conditions, p<.001. Results are illustrated in Figure 5-1, which reveals that participants spoke a significantly greater number of syllables per breath group when dictating with no ASR (17.24) than they did when dictating with either continuous-speech ASR (5.66) or discrete-word ASR (2.03). Partial eta-squared values revealed that 76% of the variance in syllables per breath group was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 5-2.

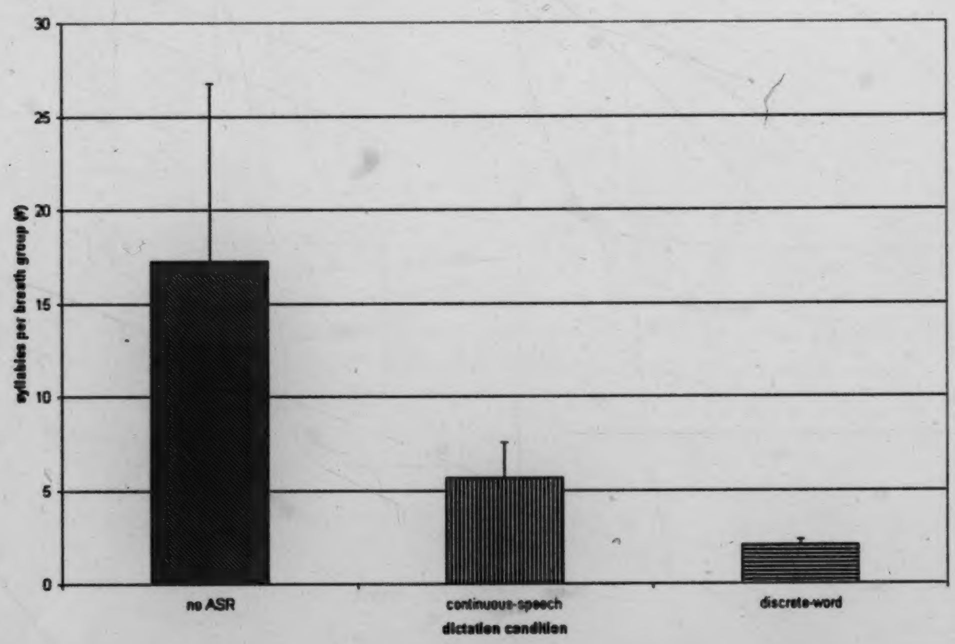

Figure 5- 1. Main effects for syllables per breath group across dictation conditions during dictation of the California Passage. Differences were significant amongst all three dictation conditions.

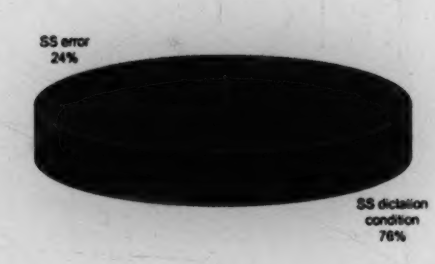

Figure 5- 2. Partial eta-squared values for syllables per breath group. The proportion of total variability in the sample attributable to dictation condition was 76%. SS represents the sum of squared differences from the mean; variance is calculated from the SS.

Frequency of Breath Groups

The findings for frequency of breath groups per reading of the California Passage are illustrated in Figure 5-3. Bonferroni post-hoc analyses revealed significant differences in frequency of breath groups amongst all three dictation conditions, p<.0001. Participants used significantly more breath groups when dictating with discrete-word software (316.25) than they did with either continuous-speech ASR (118.17) or no ASR (24.58), p<.0001. Likewise, participants used significantly more breath groups when dictating with continuous-speech ASR than when dictating with no ASR, p<.0001. Partial eta-squared values revealed that 96% of the variance in frequency of breath groups was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 5-4.

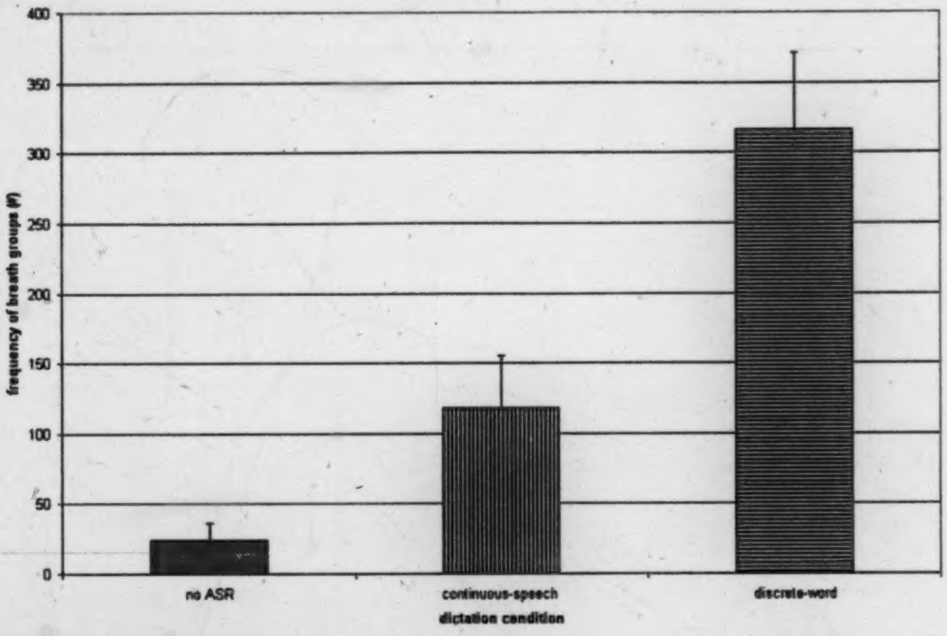

Figure 5- 3. Main effects for frequency of breath groups across dictation conditions during dictation of the California Passage. Differences were significant amongst all three dictation conditions.

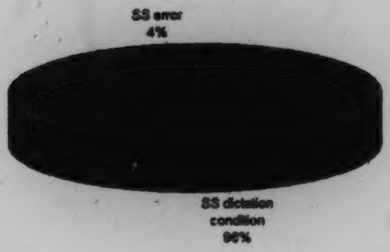

Figure 5- 4. Partial eta-squared values for frequency of breath groups. The proportion of total variability in the sample attributable to dictation condition was 96%. SS represents the sum of squared differences from the mean; variance is calculated from the SS.

Frequency of Apneic Episodes

Significant differences were found in frequency of apnea across dictation conditions during reading of the California Passage and are illustrated in Figure 5-5. Bonferroni post-hoc analyses revealed significant differences amongst all dictation conditions. There were no instances of apnea when participants dictated without ASR (0.00). This was in significant contrast to the occurrence of apnea when they were dictating with continuous-speech (7.00) and discrete-word ASR (17.08), p=.005 and p=.002, respectively. Differences between the frequency of apnea in dictation with continuous-speech ASR and discrete-word ASR also were significant at p=.012. Partial eta-squared values revealed that 66% of the variance in frequency of apnea was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 5-6.

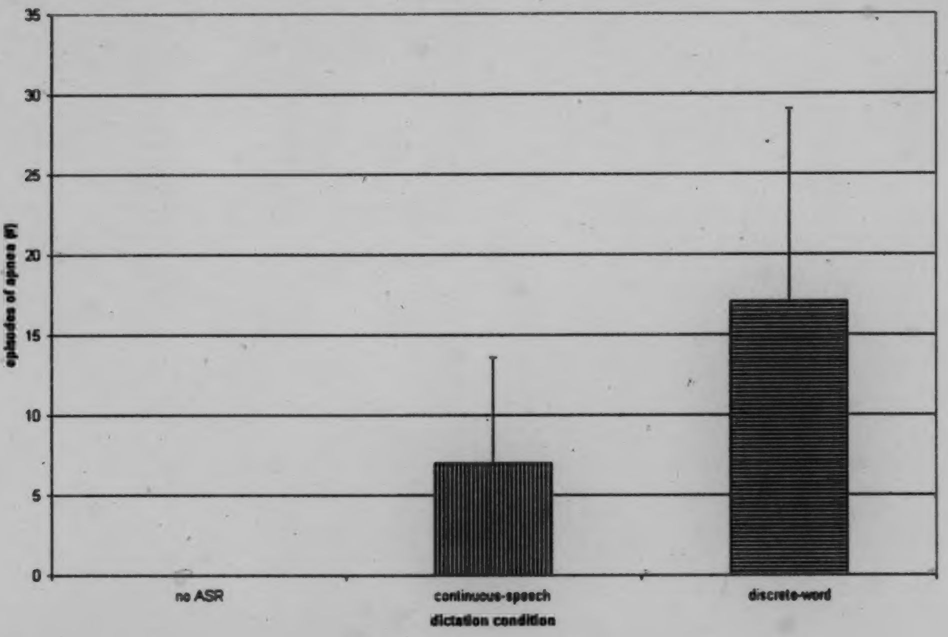

Figure 5- 5. Main effects for frequency of apnea across dictation conditions during dictation of the California Passage. Differences were significant amongst all three dictation conditions.

Figure 5- 6. Partial eta-squared values for frequency of apnea. The proportion of total variability in the sample attributable to dictation condition was 66%. SS represents the sum of squared differences from the mean; variance is calculated from the SS.

Time Needed for Dictation

Significant differences were found in the amount of time needed to dictate the California Passage across dictation conditions and are illustrated in Figure 5-7. Bonferroni post-hoc analyses revealed significant differences amongst all dictation conditions, $p<.0001$. It took significantly longer for participants to dictate the California passage using discrete-word ASR (15.46 minutes) than it did with either continuous-speech ASR (8.09 minutes) or no ASR (1.24 minutes). Partial eta-squared values revealed that 92% of the variance in time needed for dictation was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 5-8.

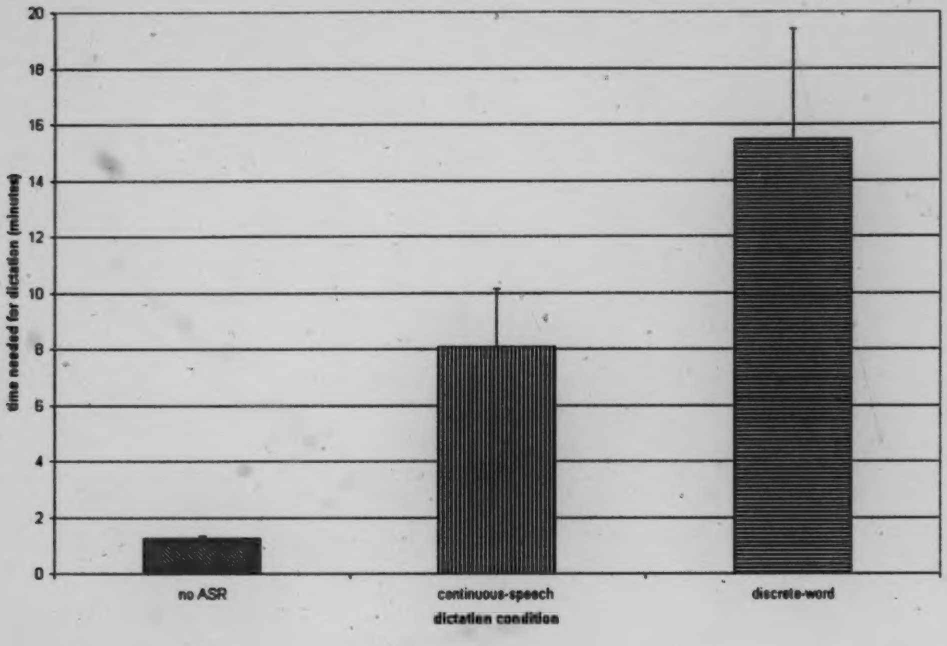

Figure 5- 7. Main effects for time needed for dictation of the California Passage across dictation conditions. Differences were significant amongst all three dictation conditions.

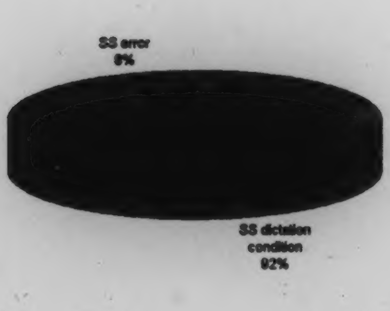

Figure 5- 8. Partial eta-squared values for time needed for dictation. The proportion of total variability in the sample attributable to dictation condition was 92%. SS represents the sum of squared differences from the mean; variance is calculated from the SS.

Number of Words Used

The number of words used in dictation of the California Passage was found to differ significantly amongst dictation conditions. The results are illustrated in Figure 5-9. Bonferroni post-hoc analyses of these results revealed significant differences between the no-ASR condition and both the discrete-word ASR and continuous-speech ASR dictation conditions, $p<.0001$. Participants used significantly fewer words when dictating without ASR (219.67) than when dictating with either continuous-speech ASR (383.67) or discrete-word ASR (427.67). There were no significant differences between the discrete-word and continuous-speech conditions, $p=.213$. Partial eta-squared values revealed that 88% of the variance in number of words spoken was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 5-10.

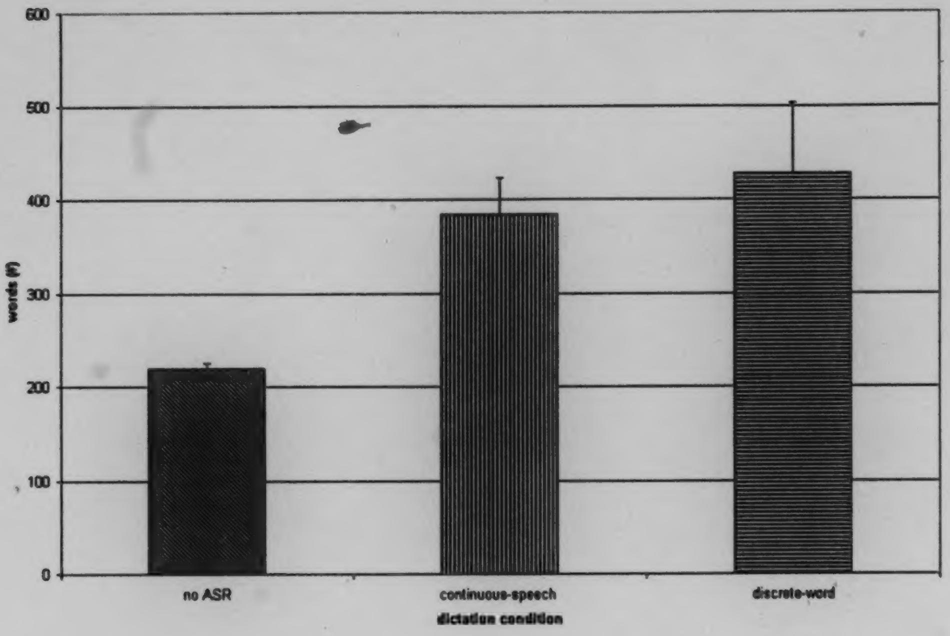

Figure 5- 9. Main effects for the number of words used in dictation of the California Passage across dictation conditions. Differences between the no ASR condition and both the continuous-speech ASR and discrete-word ASR conditions were significant. The difference between the continuous-speech ASR and the discrete-word ASR conditions was not significant.

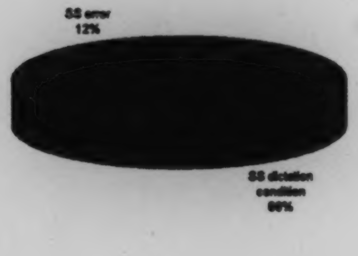

Figure 5- 10. Partial eta-squared values for number of words used in dictation of the California Passage. The proportion of total variability in the sample attributable to dictation condition was 88%. SS represents the sum of squared differences from the mean; variance is calculated from the SS.

Dictation of a Letter

Table 5-5 contains mean scores and standard deviations for the dependent variables obtained for each group across the three dictation conditions for the first three minutes of spontaneous dictation of a letter. Unlike the analyses for the California Passage, time needed for dictation was not included as a dependent variable in these analyses because the time set for spontaneous dictation of a letter in all dictation conditions was fixed at 3 minutes.

	GROUP	DICTATION CONDITION	Mean	Std. Deviation
syllables per breath group (#)	SCI	discrete-word	1.87	.18
		continuous-speech	4.65	1.45
		no ASR	10.85	3.75
	non-SCI	discrete-word	2.08	.47
		continuous-speech	4.81	.50
		no ASR	12.28	3.14
frequency of breath groups (#)	SCI	discrete-word	75.67	18.49
		continuous-speech	49.50	15.24
		no ASR	50.67	12.75
	non-SCI	discrete-word	61.33	16.91
		continuous-speech	48.33	13.92
		no ASR	47.67	12.09
frequency of apnea (#)	SCI	discrete-word	2.17	2.56
		continuous-speech	1.83	3.06
		no ASR	1.50	2.51
	non-SCI	discrete-word	1.33	1.37
		continuous-speech	2.50	2.81
		no ASR	.17	.41
number of words (#)	SCI	discrete-word	94.33	20.80
		continuous-speech	154.67	46.42
		no ASR	389.67	100.05
	non-SCI	discrete-word	87.00	12.81
		continuous-speech	151.17	38.04
		no ASR	418.33	64.27

Table 5- 5. Means and standard deviations for the speech-production variables analyzed during the first 3 minutes of dictation of a letter. Results are reported between groups and across dictation conditions.

Within-subjects results from each repeated-measures ANOVA for the speech production variables related to spontaneous dictation of a letter are summarized in Table 5-6. Between-groups results are summarized in Table 5-7. The analyses revealed significant dictation-condition main effects for syllables per breath group, frequency of breath groups, and number of words. No significant main effects were revealed for frequency of apneic episodes for dictation condition, p=.096. No interaction effects were found. In addition, no significant differences were found between speaker groups for any of the speech-production variables during spontaneous dictation of a letter. Because no group differences were found, the results that are illustrated in Figures 5-11, 5-13, and 5-15 represent main effects for dictation condition averaged across all speakers. For the reader who is interested in group means, please refer to Table 5-5.

Univariate Tests

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared	Observed Power
DICTATION CONDITION	syllables/breath group	Greenhouse-Geisser	585.341	1.153	507.610	69.032	.000*	.873	1.000
	frequency of breath groups	Sphericity Assumed	3029.389	2	1514.694	7.582	.004*	.431	.908
	frequency of apnea	Sphericity Assumed	11.167	2	5.583	2.638	.096	.209	.464
	number of words	Greenhouse-Geisser	660362.722	1.312	503281.508	99.193	.000*	.908	1.000
DICTATION CONDITION * GROUP	syllables/breath group	Greenhouse-Geisser	3.099	1.153	2.688	.366	.587	.035	.068
	frequency of breath groups	Sphericity Assumed	305.167	2	152.583	.764	.479	.071	.161
	frequency of apnea	Sphericity Assumed	6.500	2	3.250	1.535	.240	.133	.287
	number of words	Greenhouse-Geisser	2345.389	1.312	1787.434	.352	.621	.034	.069

Table 5- 6. Summary of ANOVA results for within-subjects results for each speech production variable during spontaneous dictation of a letter. *Indicates a statistically significant result.

Tests of Between-Subjects Effects

Transformed Variable: Average

Source	Measure	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared	Observed Power
GROUP	syllables/breath group	3.187	1	3.187	.664	.434	.062	.115
	frequency of breath groups	342.250	1	342.250	1.214	.296	.108	.170
	frequency of apnea	2.250	1	2.250	.190	.672	.019	.068
	number of words	318.028	1	318.028	.127	.729	.013	.062

Table 5- 7. Summary of ANOVA results for between-groups results for each speech production variable during spontaneous dictation of a letter.

Average Number of Syllables per Breath Group

Significant differences were found for the number of syllables spoken per breath group when dictating a letter. The results are illustrated in Figure 5-11. Bonferroni post-hoc analyses of the significant dictation-condition finding revealed significant differences amongst all the dictation conditions, p<.0001. Participants used a significantly greater number of syllables per breath group when dictating with no ASR (11.57) than they did when dictating with either continuous-speech ASR (4.73) or discrete-word ASR (1.97). Partial eta-squared values revealed that 87% of the variance in syllables per breath group during spontaneous speech was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 5-12.

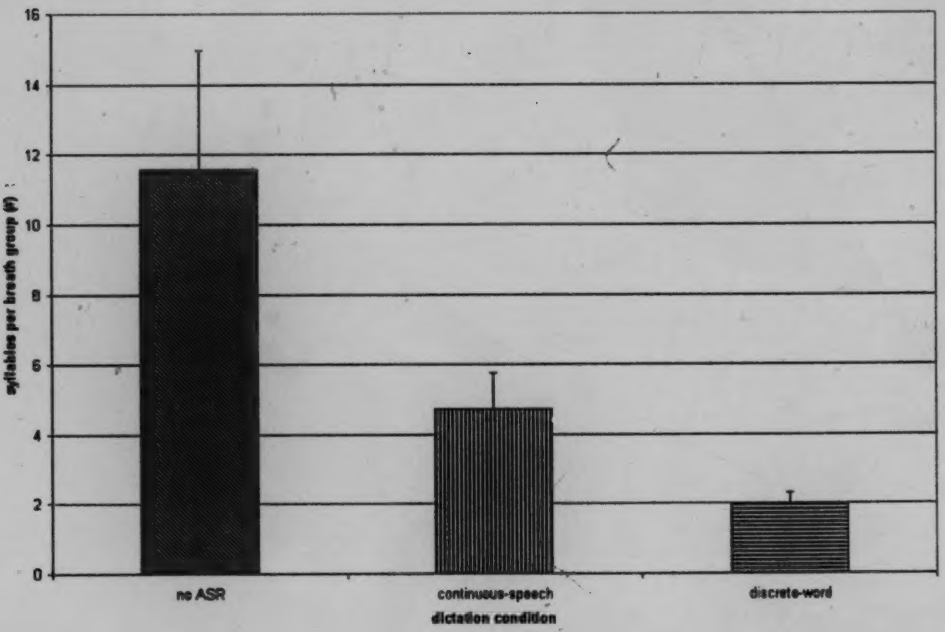

Figure 5- 11. Main effects for syllables per breath group across dictation conditions during spontaneous dictation of a letter. Differences were significant amongst all three dictation conditions.

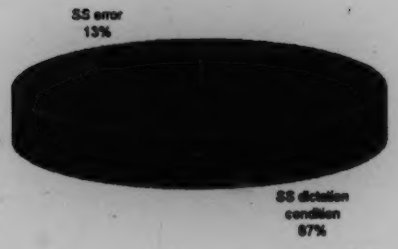

Figure 5- 12. Partial eta-squared values for syllables per breath group during dictation of a letter. The proportion of total variability in the sample attributable to dictation condition was 87%. SS represents the sum of squared differences from the mean; variance is calculated from the SS.

Frequency of Breath Groups

Significant differences were found amongst dictation conditions for the frequency of breath groups in the first three minutes of spontaneous dictation of a letter. The results are illustrated in Figure 5-13. Bonferroni post-hoc analyses revealed significant differences between dictation with discrete-word ASR and both the continuous-speech and no-ASR dictation conditions, p=.05 and p=.004, respectively. Participants used significantly more breath groups to dictate a letter when using discrete-word software (68.5) than they did with either continuous-speech ASR (48.92) or no ASR (49.17). No significant differences were found in frequency of breath groups between the continuous-speech and no-ASR dictation conditions, p>.06. Partial eta-squared values revealed that 43% of the variance in frequency of breath groups during spontaneous speech was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 5-14.

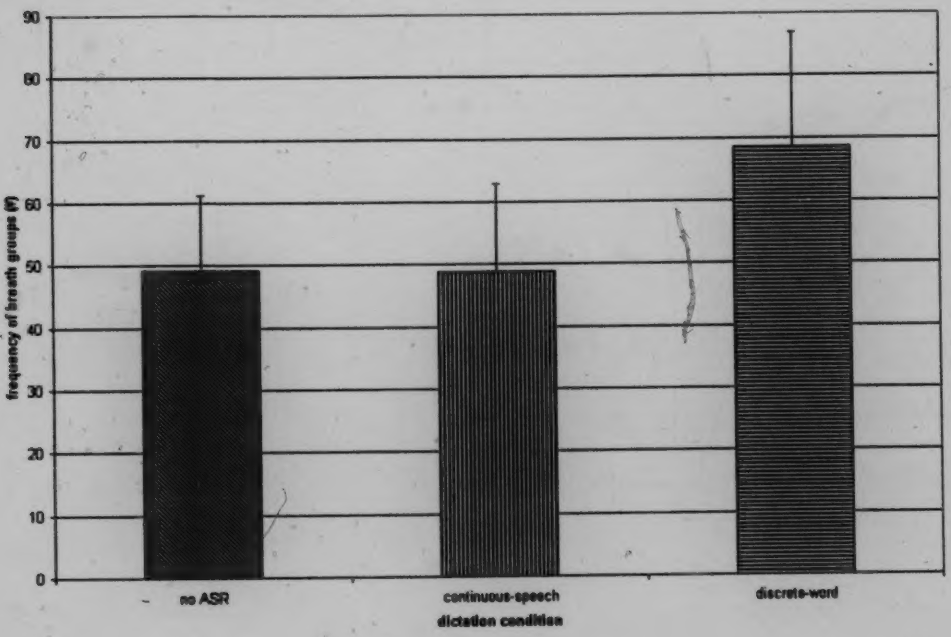

Figure 5- 13. Main effects for frequency of breath groups across dictation conditions during spontaneous dictation of a letter. Differences between the discrete-word ASR condition and both the continuous-speech ASR and no-ASR conditions were significant. There was no significant difference between the continuous-speech ASR and the no-ASR conditions.

Figure 5- 14. Partial eta-squared values for frequency of breath groups during dictation of a letter. The proportion of total variability in the sample attributable to dictation condition was 43%. SS represents the sum of squared differences from the mean; variance is calculated from the SS.

Number of Words Used

Significant differences were found amongst dictation conditions for the number of words spoken in the first three minutes of spontaneous dictation of a letter. The results are illustrated in Figure 5-15. Bonferroni post-hoc analyses revealed significant differences amongst all dictation conditions, p<.003. Participants used significantly more words when dictating with no ASR (404.00) than they did when dictating with either continuous-speech ASR (152.92) or discrete-word ASR (90.67), p<.003. Correspondingly, participants used significantly fewer words when dictating with discrete-word ASR than they did when dictating with continuous-speech ASR, p<.003. Partial eta-squared values revealed that 91% of the variance in the number of words spoken during three minutes of spontaneous speech was attributable to dictation condition. The remaining proportion of the variance was due to error. This is illustrated in Figure 5-16.

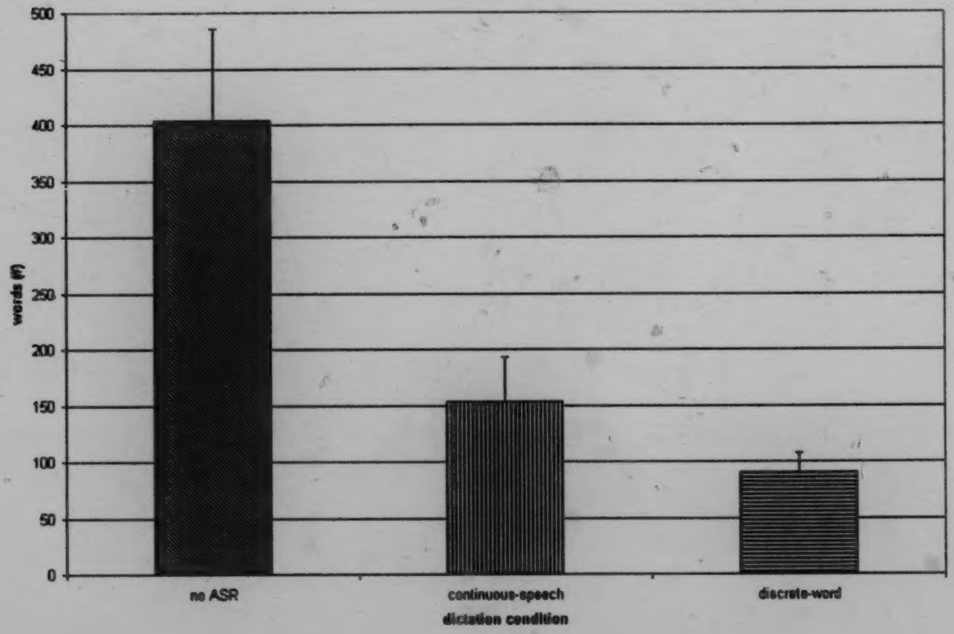

Figure 5- 15. Main effects for number of words spoken during spontaneous dictation of a letter across dictation conditions. Differences were significant amongst all three dictation conditions.

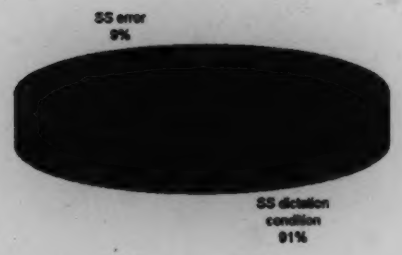

Figure 5- 16. Partial eta-squared values for number of words spoken during spontaneous dictation of a letter. The proportion of total variability in the sample attributable to dictation condition was 91%. SS represents the sum of squared differences from the mean; variance is calculated from the SS.

Measurement Reliability

Measurement reliability was assessed via intra-class correlations (ICC) for 25% of the speech production data collected in this portion of the investigation. The ICC coefficient for reliability was .9897. A confidence interval of 95% was established with a lower bound of .9888 and an upper bound of 1.0, which encompassed the ICC coefficient for reliability obtained for these repeated measurements.

Discussion

The intent of this portion of the investigation was to determine if speech-production behaviors differed when individuals dictated a written paragraph and a spontaneously generated letter with discrete-word ASR, continuous-speech ASR, and no ASR. In addition to the differences observed across these dictation conditions, of interest was the possibility that group differences may exist between individuals with SCI and their control-matched cohorts. The results of this portion of the experiment will be discussed in the text that follows with respect to dictation conditions and speaker groups, and will be compared to the outcomes of other studies of both non-disabled individuals and those with SCI reported in the literature. In addition, the implications of ASR-induced changes in speech-production behavior for users will be explored, as will the effect on a user's dictation efficiency and vocal health.

Average Number of Syllables per Breath Group

The mean number of syllables produced per breath group by the control speakers while they dictated the California Passage in the control condition was 21.27. This datum falls near the high end of ranges reported in other studies for non-disabled individuals during reading tasks (Hodge & Rochet, 1989; Hoit & Hixon, 1987; Hoit, Hixon, Altman & Morgan, 1989; Manifold & Murdoch, 1993; Solomon & Hixon, 1993) as shown in Table 5-8.

Authors	Syllables per breath group	Stimulus
Present Study	21.27	California Passage
Hoit & Hixon (1987)	16.26 - 20.77	California Passage
Hoit, Hixon, Altman & Morgan (1989)	19.85 - 20.86	California Passage
Hodge & Rochet (1989)	16.00	First paragraph of the Rainbow Passage
Solomon & Hixon (1993)	17.50	2nd & 3rd sentences of the Rainbow Passage
Manifold & Murdoch (1993)	16.25 - 19.78	Rainbow Passage

Table 5- 8. Syllables per breath group recorded in the present study and others for non-disabled individuals reading aloud under natural speaking conditions.

Syllables produced per breath group during spontaneous speech have been reported less frequently than during read speech. One group of authors (Mitchell, Hoit & Watson, 1996) reported a range of 16 to 18 syllables per breath group during spontaneous speech. In the present investigation, the mean number of syllables per breath group produced by the control group while dictating a letter in the no-ASR condition was somewhat lower (12.28 syllables). A possible explanation for the difference between the results reported by Mitchell et al. (1996) and those reported herein could be due to the structure of the speaking tasks that were used to elicit the spontaneous discourse. Subjects in the study by Mitchell and colleagues were asked to speak on topics for which an opinion had to be generated in a free-flowing monologue. In the present study, subjects were asked to dictate a letter to a friend, and the transcripts reveal that some breath groups consisted of short statements related to the form of a letter, such as "Dear Bill", "How are you?", "I am fine", "Best Regards", "Kathi". It is likely that short statements such as these influenced the means calculated for the number of syllables per breath group in the spontaneous-speech sample in the present study.

With respect to dictation with ASR, it was hypothesized that the number of syllables produced per breath group would differ amongst dictation conditions based on the predicted influence of ASR on speech-production behaviors and the results of pilot-study work. Specifically, it was predicted that dictation in the control condition would be characterized by more syllables per breath group than dictation with discrete-word ASR, and the number of syllables produced during dictation with continuous-speech ASR was predicted to fall between the other two conditions. Because discrete-word ASR requires a pause between each word that is dictated, it was expected that breath groups would be short. Short breath groups during dictation with this version of ASR were observed during a pilot study of two individuals not included in the present investigation; their breath groups in the discrete-word dictation condition consisted primarily of 1- to 2-word utterances. Because continuous-speech ASR allows more natural speech input than discrete-word ASR, breath groups produced during dictation with this software were expected to reflect juncture that more closely approximated normal speaking patterns.

Pilot-study work had revealed breath groups during dictation with continuous-speech ASR that consisted primarily of 3- to 5-word utterances.

The hypothesis developed for this dependent variable across dictation conditions was fully supported by the results for both scripted and spontaneous speech (see Figures 5-1 & 5-11). Significant differences were revealed amongst all dictation conditions, with participants speaking the greatest number of syllables per breath group during dictation in the control condition and fewer syllables per breath group when dictating with ASR. These findings suggest that discrete-word and continuous-speech ASR constrain the natural flow of speech to a degree that differs from normal. Further to this, although dictation with continuous-speech ASR constrains the flow of natural speech, it does not do so to the same extent as dictation with discrete-word ASR. Theoretically, this pattern of speech-production behavior could be related to the amount of expiratory volume available for speech, the speech rate associated with each type of dictation, or both, as will be discussed next.

The differences in number of syllables spoken per breath group between dictation with continuous-speech ASR and no ASR do not appear to be related to the available air volume for speech, as results from Chapter 3 reveal that volume expenditure per expiratory segment of speech did not differ significantly between these dictation conditions. Thus, an alternative explanation would be that speech rate might have influenced the number of syllables produced per breath group. Speech rate can be calculated in words-per-minute (WPM), and will be done so herein to illuminate differences in speech production across dictation conditions and allow for comparison of the current data to ASR manufacturers' reports of speaking rates noted in Chapter 1.

Calculation of the rate of speech production from the experimental means (Table 5-9) confirms that WPM rates were slower during dictation with continuous-speech ASR than in the control condition, which is likely to result in the production of fewer syllables per breath group for the same expiratory air volume that is available for speech.

Considering discrete-word ASR, the fact that fewer syllables were produced with that technology than in the other conditions could be due to both a smaller available expiratory volume for speech and to a slower speaking rate in that condition. A smaller expiratory air supply for speech during dictation in this condition would restrict the number of syllables that could be produced per expiratory segment, unless an individual were to increase speech rate so that a greater number of syllables could be produced. However, WPM rates were slowest in this condition (See Table 5-9). Therefore, the reduction in number of syllables spoken per breath group during dictation with discrete-word ASR could possibly be related to both restrictions in expiratory air supply and speech rate.

		Word-Per-Minute Rates	
	No ASR	Continuous-speech ASR	Discrete-word ASR
California Passage	183	47	28
Letter	134	51	30

Table 5- 9. Mean word-per-minute rates for all participants across dictation conditions for both speaking tasks.

Although significant differences in the number of syllables produced per breath group were found across all dictation conditions, no significant differences were revealed between speaker groups in the ANOVA. Hypotheses generated *a priori* for both spontaneous and scripted speech stated that syllables per breath group would be fewer for individuals with SCI than for non-disabled individuals. This hypothesis was based on research by Hoit, Banzett, Brown and Loring (1990) in which the number of syllables per breath group for individuals with SCI ranged between 6 and 15 syllables across extemporaneous speech and reading tasks, a value lower than that typically found for non-disabled individuals, which ranges between 11 and 21 syllables (Hodge & Rochet, 1989; Hoit & Hixon, 1987; Hoit et al., 1989; Manifold & Murdoch, 1993; Solomon & Hixon, 1993).

Although not statistically significant, the differences in the number of syllables produced per breath group between speaker groups in the present study supported the direction of the *a priori* hypothesis, and the data for the speakers with SCI fell within the upper portion of the range reported by Hoit et al. (1990) for their subjects. As is shown in Table 5-10, individuals with SCI produced fewer syllables per breath group than their non-disabled cohorts during dictation of both speaking tasks in the control condition.

	Syllables Per Breath Group	
	Control Subjects	Subjects with SCI
California Passage	21.3	13.2
Letter	12.3	10.9

Table 5- 10. Mean number of syllables produced per breath group by speakers in both groups during dictation with no ASR for both speaking tasks.

Although differences between the two speaking tasks were not examined within the ANOVA analyses, the data for syllables per breath group during each speaking task can be described. The only trend in differences between the two speaking conditions was that seen in the control condition. As can be observed in Figure 5-17, when subjects dictated with no ASR, they produced a greater number of syllables per breath group when reading the California Passage than they did when spontaneously dictating a letter. This is not unlike the results reported for syllables per breath group by Hodge and Rochet (1989). Their subjects produced a greater number of syllables per breath group when reading than during spontaneous speech, reportedly because oral reading is associated with reduced linguistic formulation demands and therefore, fewer pauses and interruptions in the flow of speech than those observed during spontaneous speech. Interestingly, the trend observed between speech samples during dictation in the control condition was not observed during dictation in either ASR condition. This may be due to the fact that when dictating with ASR, the linguistic demands remain high for scripted speech due to the need for formulation of corrective utterances for computer misrecognitions. In addition, the need to process computer recognition of speech input will require a greater number of pauses within a scripted dictation task, unlike the control condition where pauses and interruptions in the flow of speech will be minimal.

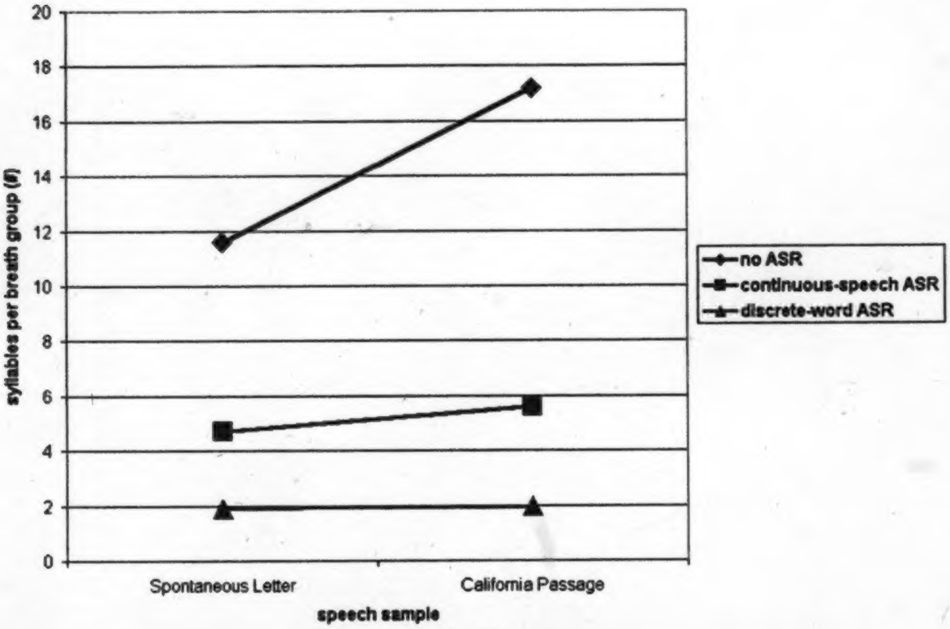

Figure 5- 17. Trends in the data for syllables produced per breath group between speech samples for all participants. Although differences were not tested statistically, the trend for dictation with no ASR reveals that more syllables per breath group are produced in a reading task than in a spontaneous speaking task. The same trend is not apparent in the plots for the discrete-word condition and only barely apparent in the continuous-speech dictation condition. [The interconnecting lines between the data points are meant to facilitate visualization of differences between the speech samples within a dictation condition for this variable.]

Frequency of Breath Groups

The frequency of breath groups for non-disabled individuals within this study in the control condition was 19.33 while reading the California Passage, and 47.66 while dictating a letter. When the number of breath groups for each speaking task was normalized to one minute, the per-minute frequency for each speaking task (16.3 for scripted speech and 15.9 for spontaneous speech) is comparable to the 16.2 breath groups per minute reported by Hodge and Rochet (1989) for both scripted and spontaneous speech.

With respect to dictation with ASR, it was hypothesized that the frequency of breath groups during dictation of the California passage and a letter would be highest when using discrete-word ASR and lowest with no ASR. This hypothesis was based on knowledge regarding rate of speech required by ASR systems for success with speech recognition, knowledge from pilot-study work, and knowledge regarding the need for spoken commands to correct misrecognitions. When speaking more slowly for the same respiratory supply, users will produce fewer words on one breath before having to re-breathe. Thus, for a passage of predetermined length, expiratory breath groups will occur more frequently when speaking at a slower rate than when speaking at a normal rate. This was confirmed in pilot-study work during dictation with both ASR systems where it was observed that more breath groups were needed to express the same utterance than were needed in the control condition. Further to this, additional breath groups were required during dictation with both ASR systems for commands spoken to the computer to correct a misrecognized utterance.

The hypothesis generated for the number of breath groups produced during dictation of the California Passage across experimental conditions was fully supported by the data collected in the current study. Significantly fewer breath groups were needed to dictate the California Passage in the control condition than in either ASR condition. With respect to the two ASR conditions, the frequency of breath groups exhibited during dictation of the California Passage with discrete-word ASR was significantly higher than the frequency exhibited during dictation with continuous-speech ASR. When considering the dictation-condition findings for frequency of breath groups along with the findings for syllables per breath group, it becomes apparent that speakers will have to produce more breath groups upon which fewer syllables are uttered in order to dictate a message with ASR, especially discrete-word ASR. This has implications for the efficiency with which a scripted-dictation task can be completed with ASR and suggests that an increased workload may be more likely with ASR systems.

Efficiency can be defined as the ability to complete a task without the waste of such things as time and energy. Further to this, it is acknowledged that the greater the proportion of energy devoted to the performance of a task, the lower the efficiency (Kroemer & Grandjean, 1997). Within this study, the control condition can be considered representative of efficient speech-production behavior. The quotient resulting from the division of efficient speech production in the control condition by the experimental conditions [continuous-speech and discrete-word ASR] yields the relative efficiency of the task in each of those conditions. Through this normalization process, a resulting relative efficiency rating of 1 would indicate that the control condition and an experimental condition are equally efficient, while a value <1 or >1 would indicate a greater or lesser degree of efficiency, respectively. The data in Table 5-11 reveal that for the number of syllables produced per breath group, the relative efficiency rating of the continuous-speech condition was 3.05, while this value was 8.50 in the discrete-word ASR condition. Thus, when compared to their efforts in the control condition, participants devoted a greater proportion of energy to the speaking task when they dictated the California Passage with ASR, resulting in a decrease in the efficiency of speech production when using the technology.

	Speech-Production Efficiency (California Passage)					
	Control		Continuous-speech ASR		Discrete-word ASR	
	Syllables/BG	Frequency BG	Syllables/BG	Frequency BG	Syllables/BG	Frequency BG
Mean	17.24	24.58	5.66	118.17	2.03	316.25
Ratio	1:1	1:1	1:3.05^{A}	1:4.81^{B}	1:8.50^{A}	1:12.87^{B}

Table 5- 11. Ratio of speech-production efficiency for one unit of work normalized to the control condition, to one unit of work under continuous-speech and discrete-word ASR conditions during dictation of the California Passage for syllables produced per breath group and frequency of breath groups. Calculation of speech-production efficiency ratio: ^{A} = CONTROL / ASR; ^{B} = ASR / CONTROL. BG = breath group.

Unlike the differences observed across conditions for dictation of scripted speech, significant differences in the number of breath groups produced amongst dictation conditions for spontaneous speech were found only between discrete-word ASR and the other two dictation conditions. When subjects spontaneously dictated a letter, use of discrete-word ASR was characterized by a higher number of breath groups than dictation with continuous-speech ASR or with no ASR. This is consistent with the finding that fewer syllables were produced per breath group when dictating with discrete-word ASR than with continuous-speech ASR or in the control condition, necessitating a higher number of breath groups when dictating with discrete-word ASR. The frequency of breath groups during spontaneous dictation with continuous-speech ASR did not differ from that recorded for the control condition (49 breath groups each). This is noteworthy in the context of results for syllable output per breath group, which was greater in the control condition than in the continuous-speech ASR condition. Thus, for the same amount of respiratory effort, as measured by the number of inspirations and expirations that were needed to dictate a letter, fewer syllables were expressed per expiration in the continuous-speech ASR condition than in the control condition. This is consistent with the slower word-per-minute rates calculated in Table 5-9 for dictation with this type of ASR and would indicate that the efficiency of the speech-production process was reduced during spontaneous dictation with continuous-speech ASR. Table 5-12 provides efficiency ratios for spontaneous speech production with and without ASR.

	Speech-Production Efficiency (Letter)					
	Control		Continuous-speech ASR		Discrete-word ASR	
	Syllables/BG	Frequency BG	Syllables/BG	Frequency BG	Syllables/BG	Frequency BG
Mean	11.56	49.17	4.73	48.92	1.97	68.50
Ratio	1:1	1:1	$1:2.44^{A}$	$1:1^{B}$	$1:5.87^{A}$	$1:1.39^{B}$

Table 5- 12. Ratio of speech-production efficiency for one unit of work normalized to the control condition, to one unit of work under continuous-speech and discrete-word ASR conditions during spontaneous dictation of a letter for syllables produced per breath group and frequency of breath groups. Calculation of speech-production efficiency ratio: ^{A} = Control condition / ASR condition; ^{B} = ASR condition / Control condition. BG = breath group.

Although differences in the frequency of breath groups were found across dictation conditions for both speaking tasks, no significant differences were revealed between speaker groups. It was hypothesized that the frequency of breath groups during both speaking tasks would be higher in individuals with SCI than in the control group. This hypothesis was based on previous research (Hoit et al., 1990) documenting that individuals with SCI produced fewer syllables per breath group than non-disabled individuals, which would result in the need for more breath groups by the speakers with SCI to relay the same passage. This assumption was not supported by the ANOVA, although the data for speakers with SCI fell in the expected direction relative to those for the non-disabled speakers. During dictation of the California Passage across all three conditions, individuals with SCI produced a higher number of breath groups overall (162) than the individuals without SCI (143). This difference was in the same direction but not as large for dictation of a letter (59 breath groups for individuals with SCI versus 52 breath groups for their non-disabled cohorts).

Comparison of the frequencies of breath groups between the two speech samples was not completed statistically because of the inherent differences in the tasks that are related to the amount of time that subjects spent dictating. When comparable sampling periods of one minute are considered, however, as shown in Table 5-13, the number of breath groups produced during each task does not appear to differ substantially.

	Breath Groups Per Minute		
	No ASR	Continuous-speech ASR	Discrete-word ASR
California Passage	16	15	18
Letter	16	16	20

Table 5- 13. Mean breath groups per minute for control subjects across dictation conditions for both speech samples.

Frequency of Apneic Episodes

In the present investigation, an apneic episode was a moment in the respiratory cycle when the breath was "held". Although other investigators have described the presence of apnea in individuals with polio and Rett Syndrome (Hixon, Putnam & Sharp, 1983; Percy, Zoghbi, Lewis & Jankovic, 1988), the primary investigator is unaware of studies that have reported the frequency of apnea during speech breathing in individuals with SCI. Some investigators have alluded to the presence of apnea in the speech of non-disabled individuals (Hodge & Rochet, 1989; Mitchell, Hoit & Watson, 1993), however, the frequency with which it occurred was not reported. Thus, comparison of the frequency of apnea data during the speaking tasks within this study to other reference data is not possible.

Within the current investigation, it was hypothesized that the frequency of apnea would be highest during dictation with discrete-word ASR and lowest for dictation in the control condition. When speakers were using discrete-word technology, it was anticipated that they might not want to re-breathe between words, which are necessarily separated by pauses. Instead, it was predicted that individuals would have a tendency to hold their breath during some of these pauses. Further to this, for both ASR systems, the linguistic formulation demands that accompany dictation are assumed to be higher due to the user's need to formulate commands that direct the software to correct misrecognized utterances. Speech with higher linguistic demands may be characterized by an increased number of pauses (Hodge & Rochet, 1989; Mitchell, Hoit & Watson, 1996), and these pauses may be characterized by either a quiet expiration or a breath hold (Mitchell, Hoit & Watson, 1996). Of interest to this study were pauses that were characterized by a breath hold (i.e., an instance of apnea).

Statistical analyses revealed that expected differences amongst all dictation conditions for instances of apnea were significant during dictation of the California Passage. Specifically, dictation in the control condition was characterized by no instances of apnea, while dictation with both ASR systems revealed significant instances of apnea. Further, dictation with discrete-word ASR was characterized by a significantly greater frequency of apnea than dictation with continuous-speech ASR. The results for apnea during spontaneous dictation did not follow the same pattern. There were no significant differences found amongst any of the dictation conditions with respect to the number of breath holds that occurred during three minutes of spontaneous dictation. Results for the frequency of apnea across dictation conditions were expected to be similar for each speaking task and, therefore, the lack of correspondence between the findings for each speech sample deserves some attention.

The most likely explanation for the lack of significant differences in the frequency of apnea across dictation conditions for the letter may be related to the nature of the speaking task, to the amount of time that was spent in dictation, or both. The nature of the spontaneous-speaking task was such that if the computer did not recognize a particular word or phrase, subjects were free to vary the linguistic content of their message to something that was more easily recognized by the ASR system, or to introduce a different message altogether. The primary investigator observed this pattern of behavior for many of the subjects during spontaneous dictation of a letter. Had the freedom for linguistic variation been removed, such as when the subjects dictated the California Passage, repetitions of words or phrases would be required until recognition was successful. Theoretically, this could provoke periods of apnea during which corrective commands would be formulated to satisfy the regulatory effects of the ASR technology in the communication dyad.

Alternatively, it is reasonable to suggest that with an extended amount of time, such as that observed when subjects dictated the California Passage with either ASR system, there would be more opportunity for instances of apnea to occur. This explanation is confirmed by adjusting the sampling period so that the occurrence of apnea is calculated as a per-minute value, as shown in Table 5-14. Observation of the data in this table reveals little difference in the frequency of apnea between the two speaking tasks when it is calculated as a per-minute value, indicating that the sampling moment is an important consideration when considering the frequency of apnea. Thus, it is reasonable to suggest that if speakers were given an extended amount of time to dictate a letter, more instances of apnea could have occurred.

	Frequency of Apnea per Minute		
	No ASR	Continuous-speech ASR	Discrete-word ASR
California Passage	0	0.88	1.13
Letter	0.28	0.72	0.58

Table 5- 14. Mean frequency of apnea per minute for all participants across dictation conditions for both speaking tasks.

Although some dictation-condition differences were observed within the frequency-of-apnea data, there were no significant between-group differences. It had been hypothesized that the frequency of apnea during dictation of both speaking tasks would be higher in individuals with SCI than in the control group. This was based on knowledge regarding the constraints imposed upon the respiratory system in individuals with SCI, and previous reports of breath-holding as a compensatory strategy in an individual with an impaired respiratory system secondary to paralysis of the rib cage, diaphragm and abdomen (Hixon, Putnam & Sharp, 1983). One of the consequences of cervical SCI is paralysis of the external intercostal muscles of the rib cage. During expiration for speech in non-disabled individuals, the external intercostal muscles 'brake' the compression of the chest wall, counteracting the elastic recoil of the system at higher lung volumes and preventing expiratory air from 'rushing out' of the respiratory system. However, individuals with cervical SCI cannot rely on the external intercostals to counteract elastic recoil of the respiratory system at high lung volumes and therefore, may compensate with more than the usual amount of laryngeal airway resistance to control expiratory flow of air during a segment containing speech. This entails increasing the medial compression between the vocal folds to a greater degree than during normal voice production and possibly impeding airflow altogether to conserve the available respiratory supply for speech during interword intervals and pauses.

Notwithstanding the expectation that the presence of SCI might provoke an increase in the frequency of apnea among the speakers in the experimental group, a significant difference in instances of apnea between speaker groups was not revealed by the ANOVA for dictation of either the California Passage or a letter. However, trends in the data were obvious in the direction of this hypothesis when ASR was part of the dictation constraints. For example, as shown in Figure 5-18a, during dictation of the California Passage in the control condition, no instances of apnea were observed in either the individuals with SCI or in the control group. However, during dictation with both ASR systems, individuals with SCI exhibited a greater number of instances of apnea than the control group. This trend is not observed in Figure 5-18a for dictation of the letter where the frequency of apnea is similar for both groups in all three conditions. Thus, it appears group dissimilarities were evident only during dictation of scripted material using ASR.

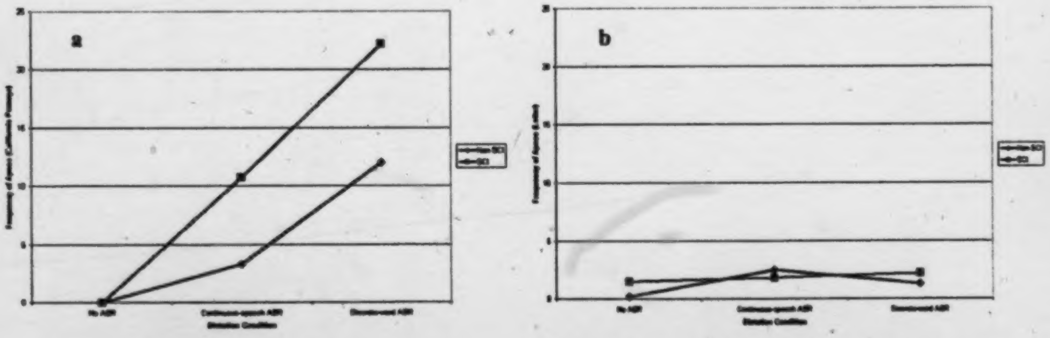

Figure 5- 18 a - b. Figure 4-18a (left) represents the frequency of apnea during dictation of the California Passage for both groups, where trends for group differences were apparent. Figure 4-18b (right) represents the frequency of apnea during dictation of a letter for both groups, where there were no trends for group differences.

The trends observed for group differences between speech tasks for frequency of apnea may be most sensibly discussed for the California Passage, which was deliberately constructed to elicit a large number of breath groups that would vary in syllable number (Hoit & Hixon, 1987). In addition, the passage contains a higher proportion of multisyllabic words than that reported for typical adult speech (Hodge & Rochet, 1989). Thus, as suggested by Hodge & Rochet (1989), a passage such as this may probe the optimal performance of the respiratory mechanism for speech by placing it under more demanding conditions, such as production of a higher proportion of longer sentences, production of a larger percentage of phonemes associated with high airflows, and production of more multisyllabic words than what may be found in spontaneous speech. In spontaneous speech, it is possible that individuals with respiratory limitations avoid such demanding aspects of speech by altering the content of their message. For example, they may become more frugal at the linguistic level by shortening the duration of consonants that are characterized by high airflow or by changing phonemes that are intense consumers of mechanical energy (Hixon, Putnam & Sharp, 1983). By making such compensations, individuals with respiratory impairments can maintain adequate respiratory support to generate the intended linguistic message, thereby reducing or avoiding the need to use the larynx in an unnatural regulatory manner. Because the trends for group differences during dictation of a scripted passage were only apparent when ASR was part of the servomechanistic loop, the suggestion is that the combination of the speaking demands associated with the California Passage plus the recognition demands of the ASR systems appear to prompt speakers, especially those with SCI, into breath-holding behavior. In spite of the group trends described above for frequency of apnea, the between-group differences were non-significant.

Time needed for Dictation of the California Passage

The mean time needed to dictate the California Passage by the non-disabled individuals in the control condition was 1 minute 20 seconds. The primary investigator is not aware of other studies that have measured mean time needed to dictate this passage during a normal speaking situation and therefore, comparison to reference data is not possible. Within the present investigation, comparisons across dictation conditions and between speaker groups revealed significant differences amongst conditions, but not between groups.

With respect to dictation condition, it was hypothesized that the time needed to dictate the California Passage would be shortest in the control condition, longest in the discrete-word condition, and at a point between the other two conditions for dictation with continuous-speech ASR. This hypothesis was based on expectations for slower word-per-minute rates when individuals dictate with ASR and on expectations that dictation time in the ASR conditions would be increased by a speaker's need to produce spoken commands associated with correcting misrecognitions by the computer.

The actual differences in time recorded for dictation amongst all dictation conditions was significant, and confirms the aforementioned expectations. Longer dictation times were consistent with the production of fewer syllables per breath group during dictation with both ASR systems, and the slower speaking rates in those conditions were confirmed by calculation of word-per-minute values for the data (See Table 5-9.). Because efficiency is related to the production of a task without waste of time, it appears that dictation with both ASR systems is less efficient than if one were to dictate naturally into a tape recorder. When considering the two ASR systems, time-efficiency will be significantly increased if continuous-speech ASR is used.

With respect to speaker groups, it was hypothesized that the time needed to dictate the California Passage would be longer for individuals with SCI than for individuals in the control group. This was based on a report that speakers with SCI produced fewer syllables per breath group than non-disabled individuals (Hoit et al, 1990) and therefore, might require a higher number of breath groups to dictate the passage. An increase in the number of breath groups needed to dictate the passage would also result in an increase in the number of inspirations to support these breath groups, with each inspiration characterized by its own inspiratory pause time, which may be longer for individuals with SCI. In addition, if the frequency of apnea were found to be higher in individuals with SCI, the length of time needed for dictation would be inflated by the time spent in moments of apnea. These factors were not found to differ significantly between the speaker groups in this study, however, and the hypothesis developed for the time needed to dictate the California Passage was not supported. It appears that, irrespective of the anticipated respiratory limitations of individuals with SCI, the experimental subjects in this study were as time-efficient when dictating, with or without ASR technology, as the non-disabled controls.

Number of Words Used

Also related to efficiency is the number of words used during dictation. In the control condition, the number of words spoken during dictation of the California Passage was the number of printed words (220) minus any words that an individual may have skipped, or plus any words that may have been added such as repetition of a word within the passage. In the ASR conditions, the number of words spoken also included the commands that participants spoke to the computer to correct misrecognitions. The number of words recorded during spontaneous dictation of a letter was the total number of words spoken in three minutes.

It was hypothesized that the number of words spoken during both speech tasks would differ across dictation conditions. Specifically, for the California Passage, it was expected that the number of words spoken would be greater during dictation with ASR (due to the speaker commands associated with correcting misrecognitions) than during dictation in the control condition. The hypothesis was supported by the data collected. Interestingly, however, there was no significant difference in the number of words produced between the two ASR systems. Fewer words might be expected with continuous-speech ASR because it has a built-in grammar to enhance recognition of a word sequence. This is not unlike recognition by human listeners, which is enhanced through the use of context from surrounding utterances (Dubno, Ahlstrom & Horwitz, 2000; Flynn & Dowell, 1999). The lack of a significant difference in the number of words needed for dictation of the California Passage between the two ASR systems suggests that even though continuous-speech ASR has made technical gains in allowing a user to dictate more naturally, as is evidenced by the production of more syllables in fewer breath groups over a shorter period of time, the number of words needed for corrective utterances has not decreased significantly from what is needed during dictation with discrete-word ASR. Table 5-15 illustrates that 43% of the words dictated with continuous-speech ASR (164 of 384) and 48% of those with discrete-word ASR (207 of

427) were corrective in nature. Thus, for both ASR systems, a large proportion of a participant's energy went into producing speech that was corrective rather than specific to the message dictated, thus lowering the overall efficiency with which the task was completed when using ASR.

	Average Number of Words Spoken during Dictation of the California Passage	
	Continuous-speech ASR	Discrete-word ASR
Words related to message	220	220
Words related to corrections	164	207

Table 5- 15. Average number of words spoken, related to both the message and to corrective utterances for misrecognitions, during dictation of the California Passage for all participants.

For spontaneous dictation of a letter, it was expected that the number of words spoken across dictation conditions would reflect the limiting influence of ASR systems on message output compared to natural spontaneous speech, due to constraints on speech rate for accurate recognition. Significant differences in the number of words spoken were found amongst all three dictation conditions: word output was highest in the control condition and lowest for the discrete-word condition. When the efficiencies of the spontaneous dictation conditions are considered in Table 5-16, it can be seen that as many as 43% of the words dictated with continuous-speech ASR (66 of 153) and 54% of those with discrete-word ASR (49 of 91) were corrective in nature. Thus, for this speech sample, too, a substantial proportion of time and energy was invested in work that was not directly related to the message, especially when using discrete-word ASR.

	Average Number of Words Spoken per 3 Minutes	
	Continuous-speech ASR	Discrete-word ASR
Words related to message	87	42
Words related to corrections	66	49

Table 5- 16. Average number of words spoken, related to both the message and to corrective utterances for misrecognitions, during 3 minutes of spontaneous dictation for all participants.

The number of words spoken when dictating the California Passage and spontaneous letter was not expected to differ between individuals with and without SCI. This hypothesis was fully supported by the results of the present investigation. The absence of significant differences between speaker groups for the number of words produced during dictation of the California Passage or a letter is consistent with the fact that there were also no significant differences in the number of breath groups or the number of syllables produced per breath group between individuals with and without SCI. Thus, for the dictation tasks completed within this study, it appears that individuals with SCI were as efficient when dictating a message as the control subjects.

Speech-Production Data Concluding Remarks

For both dictation of the California Passage and of an extemporaneous letter, differences in participants' speech-production behaviors were most obvious amongst dictation conditions. Some trends towards differences were observed between speaker groups during dictation of the California Passage, but no significant differences were found, nor were trends or significant differences apparent between speaker groups for spontaneous dictation of a letter.

With respect to speaker groups, there were no significant differences between individuals with and without SCI, even though there were trends in the data in the direction of differences that had been hypothesized. One of the primary concerns with respect to lack of significant findings between groups was the lower statistical power associated with all of the between-group analyses. With a larger subject sample size, it is possible that significant differences on some variables might emerge between individuals with and without SCI.

The observed differences among dictation conditions suggests that discrete-word ASR perturbed the speech-production mechanism from its normal behavior during speech and constrained it to a greater degree than continuous-speech ASR by reducing the number of syllables produced per breath group, increasing the frequency of breath groups, and increasing the frequency of apnea. Continuous-speech ASR also produced a change in these speech-production behaviors to a degree that differed from normal, but the change was not as great as that seen for discrete-word ASR. In addition, when participants were dictating a written passage with either type of ASR, the amount of time and the number of words they used were significantly greater than the time and words associated with production of the same message without ASR, requiring what might be considered "more work" on the part of the speaker and ultimately reducing the efficiency with which a message was produced.

As discussed in chapter 1, Baber and colleagues (1996) have defined workload in the context of speech production from a mechanical perspective, stating that an increased workload produces measurable changes in speech production. The changes in the speech-production variables measured during the use of ASR in this study suggest an increased workload imposed upon the speech apparatus. This holds true for both discrete-word and continuous-speech ASR, though less so for the latter. Reduction of efficiency was confirmed during dictation with ASR by the large proportion of words, and therefore more time and energy, that was spent in dictation of corrective utterances instead of words representing the content of the speaker's message. In addition, that efficiency was compromised during dictation with both ASR systems for both speaking tasks was revealed in lower word-per-minute rates than those observed in the control condition. Finally, while continuous-speech ASR systems appear to involve a smaller workload and are more efficient than discrete-word systems, they still involve a greater workload and are less efficient than natural speech.

Of concern to this study was the effect that ASR systems may have on the voice. From the results of this portion of the investigation with respect to workload and efficiency, it becomes apparent that ASR systems have the potential to put a speaker in a situation where they are at risk for vocal overuse and possibly fatigue. As was discussed in the introduction to this body of work, overuse of the voice is related to speaking for an extended period of time and may result in fatigue of the laryngeal system. In consideration of the efficiency issues reported for the use ASR, it can be argued that speakers may encounter situations during dictation with the technology where up to half of the words spoken are not directly related to the message they wish to convey, which will ultimately lead to increases in speaking time and workload. Further, this situation is likely to be associated with unnatural speech-production behaviors such as increased frequency of breathing for speech and slower speaking rates. Additionally, issues related to speech-breathing behaviors that were discussed in Chapter 3, such as the respiratory volumes available for speech and the extent to which those volumes are used for speech production, need to be considered in the overall evaluation of the effect of this technology on a speaker regardless of that person's neurological status. These issues will be discussed further in the concluding chapter of this manuscript.

CHAPTER 6 – SUMMARY AND CONCLUSIONS

Summary of Findings

The present investigation was an initial exploration of the nature of speech-breathing, kinematic and speech-production behaviors associated with dictation using ASR software in individuals with and without spinal cord injury. The following summary will focus first on the relative changes observed in the dependent variables as a function of dictation condition. Then, because the changes in one dependent variable relative to another provide insight into the effects of ASR on the speaking mechanism, the relationships that exist between the speech-breathing and speech-production variables will be explored. Finally, after a review of these interrelationships, discussion will focus on the influence of ASR on the speech-breathing, speech-production and kinematic variables with respect to workload, efficiency, and fatigue issues.

Regulation of Speech Without ASR

When ASR was not part of the speech servomechanism, speech-breathing and speech-production behaviors of individuals with and without SCI were regulated by the natural aeromechanical, acoustic, tactile and kinesthetic properties of connected speech. It is recognized that speech also may have been regulated by the task of speaking into a tape recorder, which may differ from speech produced in a conversational dyad. When dictating with no ASR, the participants in this study exchanged approximately 16% of their vital capacity during respiratory cycles that involved speech. Inspiration for speech lasted roughly half a second. On each expiratory cycle for speech, approximately half a litre of air was expended and, of this, about 90% was used to produce 12-17 syllables of speech. The frequency of breath groups depended on the speaking task, but averaged between 16 and 20 breath groups per minute. The number of words produced also depended on the speaking task, but averaged 134 and 183 words per minute for spontaneous and scripted speech, respectively. These speech-breathing and speech-production behaviors allowed for speedy dictation of the written passage and for a substantial letter to be dictated. Apnea was rare during natural dictation. Finally, it appears that most of the speech-breathing data for non-disabled individuals was produced with muscular pressure from both the rib cage and abdomen, but with the rib cage predominating. Chest wall part contribution for individuals with SCI also revealed a high percentage of rib cage contribution to lung volume exchange. For the most part, background configuration of chest wall revealed that most speakers, even some with SCI, assumed the shape of an inverted pear (Hixon, Goldman & Mead, 1973) during natural dictation, such that the rib cage was larger and the abdomen was smaller than their relaxed size at the same lung volume. All of the aforementioned speech behaviors can be considered to be naturally regulated through the servomechanistic loop that operates during connected speech.

ASR and the Regulation of Speech

When participants dictated with continuous-speech ASR, several changes from natural regulation were detected. As can be observed in Table 6-1, there were no significant changes in respiratory volumes exchanged for speech, however more respiratory pump cycles were observed when using continuous-speech ASR, and apnea occurred more frequently. Interestingly, the duty cycle was much lower in the continuous-speech ASR condition than in the control condition, and this is consistent with the finding that the number of syllables produced per breath group was significantly lower in the continuous-speech condition. Dictation was slower with continuous-speech ASR, and a large proportion of the message consisted of corrective utterances. Thus, it appears that the amount of energy invested in the speaking task was higher (as evidenced by an increase in the frequency of breath groups, apnea, and number of words required to dictate a message), but the use of the available air supply was less efficient (as evidenced by a decreased duty cycle, a reduction in the number of syllables spoken per breath group when the same respiratory supply was available, and an increase in the time needed to dictate a message) for dictation with continuous-speech ASR than for natural dictation. The kinematic results revealed that percent rib cage contribution to lung volume exchange during dictation with continuous-speech ASR was slightly higher than during dictation with no ASR. In addition, levels for lung volume initiation during speech were higher in both groups when they used continuous-speech ASR than when they dictated naturally. Lung volume termination levels were higher for non-disabled individuals when they used continuous-speech ASR than when they dictated with no ASR; this was not apparent in the group of individuals with SCI. Finally, background configurations of the chest wall system did not appear to differ between dictation with continuous-speech ASR and no ASR.

Continuous-speech ASR versus no ASR		
Variable	**Change?**	**Direction of Change**
% VC inspired	No	
% VC expired	No	
Mean volume expired	No	
Duty cycle	Yes	Less speech per expiration for CS-ASR
Inspiratory pause time	Yes	Longer for CS-ASR
Syllables per breath group	Yes	Fewer syllables produced with CS-ASR
Frequency of breath groups	Yes	Higher during scripted dictation with CS-ASR
Frequency of apnea	Yes	More often during scripted dictation with CS-ASR
Time needed for dictation	Yes	Longer when dictating with CS-ASR
Number of words spoken	Yes	More for scripted dictation and less for spontaneous dictation with CS-ASR
%RC contribution	Yes	Slightly higher with CS-ASR
Lung volume initiation/termination	Yes	Initiation and termination higher with CS-ASR
Background configuration	No	

Table 6- 1. Changes in speech-breathing and speech-production behaviors during dictation with continuous-speech ASR as compared to dictation with no ASR. CS-ASR = continuous-speech ASR.

When participants dictated with discrete-word ASR, changes from the control condition were detected for all dependent variables. As shown in Table 6-2, participants exhibited significant decreases in the amount of available respiratory volume for speech, more respiratory pump cycles, and more instances of apnea. Duty cycle was lower, and the number of syllables produced per breath group was much lower in the discrete-word ASR condition than in the control condition. Dictation was slow with discrete-word ASR, and a large proportion of the message consisted of corrective utterances. Thus, it appears that speech breathing and speech production are far less efficient when speakers use discrete-word ASR than when they speak naturally. The kinematic results revealed that percent rib cage contribution to lung volume exchange during dictation with discrete-word ASR was slightly higher during scripted dictation than when participants read aloud naturally, but did not differ during spontaneous dictation. Lung volume initiation levels for speech did not differ between dictation with discrete-word ASR and no ASR however, lung volume termination levels were lower during natural dictation. Background configurations of the chest wall system did not differ.

Discrete-word ASR versus no ASR		
Variable	Change?	Direction of Change
% VC inspired	Yes	Less for DW-ASR
% VC expired	Yes	Less for DW-ASR
Mean volume expired	Yes	Less for DW-ASR
Duty cycle	Yes	Less speech per expiration for DW-ASR
Inspiratory pause time	Yes	Longer for DW-ASR
Syllables per breath group	Yes	Fewer syllables produced with DW-ASR
Frequency of breath groups	Yes	Higher with DW-ASR
Frequency of apnea	Yes	More often during scripted dictation with DW-ASR
Time needed for dictation	Yes	Longer when dictating with DW-ASR
Number of words spoken	Yes	More for scripted dictation and less for spontaneous dictation with DW-ASR
%RC contribution	Yes	Slightly higher with DW-ASR during scripted dictation
Lung volume initiation/termination	Yes	Lung volume termination higher for DW-ASR
Background configuration	No	

Table 6- 2. Changes in speech-breathing and speech-production behaviors during dictation with discrete-word ASR as compared to dictation with no ASR. DW-ASR = discrete-word ASR.

When the two ASR systems are compared, it becomes obvious that continuous-speech ASR has made some gains over discrete-word ASR in more closely approximating a natural speaking condition. Continuous-speech ASR allowed for exchange of respiratory volumes of air for speech that approximated normal and therefore, more words could be spoken on fewer respiratory pump cycles than what was observed for discrete-word ASR. Apnea also occurred less frequently with continuous-speech ASR. The number of syllables produced per breath group was significantly greater in the continuous-speech condition, which is consistent with a faster speech rate and a larger available respiratory supply for speech output in that condition. However, lack of difference in duty cycle between the two conditions suggests that adjustments in the respiratory volumes available for speech were made in proportion to the anticipated amount of speech that would be expressed on one breath, and that a similar proportion of potential respiratory air supply for speech is 'wasted' during dictation with both ASR systems. As mentioned, dictation was faster with continuous-speech ASR and therefore, more words could be spoken within a set period of time during spontaneous dictation. However, an equally large proportion of the spontaneously dictated message consisted of corrective utterances when using continuous-speech ASR and discrete-word ASR. Therefore, improvements in continuous-speech ASR have not necessarily addressed the problem of computer misrecogntion of speech input. Thus, although continuous-speech ASR more closely approximates natural dictating behaviors than discrete-word ASR, issues surrounding the increased speaking workload and the efficiency with which a message can be relayed with either system still exist. Finally, the kinematic results revealed that percent rib cage contribution to lung volume exchange was slightly higher during dictation of a letter with continuous-speech ASR than it was when dictating with discrete-word ASR. However, percent rib cage contribution did not differ between the two ASR conditions during scripted dictation. Lung volume initiation levels for speech were higher and termination levels lower when individuals used continuous-speech ASR than when they used discrete-word ASR. Background configurations of the chest wall system did not appear to differ between the two ASR dictation conditions.

Discrete-word ASR versus continuous-speech ASR		
Variable	Change?	Direction of Change
% VC inspired	Yes	Less for DW-ASR
% VC expired	Yes	Less for DW-ASR
Mean volume expired	Yes	Less for DW-ASR
Duty cycle	No	
Inspiratory pause time	No	
Syllables per breath group	Yes	Fewer syllables produced with DW-ASR
Frequency of breath groups	Yes	Higher with DW-ASR
Frequency of apnea	Yes	More often during scripted dictation with DW-ASR
Time needed for dictation	Yes	Longer when dictating with DW-ASR
Number of words spoken	Yes	Less for spontaneous dictation with DW-ASR
%RC contribution	Yes	Slightly higher with CS-ASR during spontaneous dictation
Lung volume initiation/termination	Yes	Initiation levels higher; termination levels lower
Background configuration	No	

Table 6- 3. Changes in speech-breathing and speech-production behaviors when participants dictated with discrete-word ASR as compared to dictation with continuous-speech ASR. DW-ASR = discrete-word ASR; CS-ASR = continuous-speech ASR.

Many of the speech-breathing and speech-production variables that have just been described are logically related. Changes in one relative to another provide insight into those interrelationships as well as to the effects of ASR on the speaking mechanism. With respect to the speech-breathing variables, the percent vital capacity that is expired for speech is highly related to the percent vital capacity inspired on the same respiratory pump cycle. This is confirmed within this study through observation that these two variables were characterized by means that were the same regardless of speaker group, dictation condition, or speech sample. Thus, it appears that the respiratory pump worked in a consistent inspiratory-expiratory exchange mode regardless of speaking demands or physical limitations imposed upon the speakers within this study. Inspiratory pause time was initially predicted to be related to the depth of inspirations for speech, however it was found that inspiratory pause times were inversely related to inspiratory volumes. Thus, it appears that inspiratory pause time was influenced by other factors within the experimental design such as dictation condition, the influence of which will be discussed shortly.

A number of interrelationships can be noted amongst the speech-production variables. For example, the number of syllables that are produced per breath group is reciprocally related to the frequency of breath groups needed to dictate a message. Together, these two variables may influence the total speaking time because an increase in the number of breath groups needed to dictate a message would also result in an increase in the number of inspirations to support these breath groups, with each inspiration characterized by its own inspiratory pause time. At a constant speaking rate, the number of words spoken during dictation will also affect total speaking time, with the production of more words requiring more dictation time.

Finally, several interrelationships can be noted between some of the speech-production and speech-breathing variables. The depth of inspiration and expiration for the production of speech will have a bearing on the number of syllables that can be produced on the expiration. Small respiratory volumes will support fewer syllables than larger respiratory volumes, as long as speaking rate is kept constant. Therefore, smaller expiratory volumes for speech will require more inspirations and expirations in order to dictate a message. The duty cycle of the respiratory system is also related to the number of syllables that are produced per breath group, with fewer syllables resulting in lower duty cycle values and wasting of expiratory air that could be used to support speech. Theoretically, when duty cycle is low, speakers will need to increase the frequency of respiratory pump cycles to dictate a message. However, if the average volume of air inspired is adjusted (i.e., decreased) for the intended production of fewer syllables, then duty cycle values should remain high.

The natural interrelationships among the speech-breathing and speech-production variables reveal that the speaking apparatus assumes an increased workload and that efficiency of dictation is affected when ASR becomes part of an individual's speech servomechanism. The issues of workload and efficiency will be discussed next and will be followed by discussion of the relationship between these issues and development of vocal fatigue.

ASR and the Speaking Mechanism

Workload and Efficiency

From the results revealed for the variables within this study, it becomes apparent that ASR systems have the potential to put speakers in a situation where they are at risk for increased workloads and decreased efficiency of speech production, and thus potential overuse of the speech-production system. Larger-than-normal workloads during dictation with ASR were revealed in increases in the frequency of respiratory pump cycles, occurrences of apnea, and the number of words produced to dictate a message. Such increased workloads have implications for the amount of muscular energy that is expended during the dictation task. The muscles of the respiratory system will be required to expend more energy to support more inspiratory and expiratory exchanges over the course of the dictation task with ASR. In addition, during periods of apnea, individuals may recruit the diaphragm or external intercostal muscles to 'halt' the respiratory system, which will impose energy demands on those muscles. Alternatively, volume displacement from the respiratory system may be halted during periods of apnea by muscular action within the laryngeal system. Adjustments of the muscles of the larynx made in an effort to close the airway momentarily so that air cannot escape will impose energy demands on those muscles. Finally, to support the production of more words, muscular energy of the respiratory, laryngeal, and articulatory systems will be needed, over and above what is required when dictating without ASR. Thus, when considering the dictation task as a whole, it is apparent that an increased amount of energy goes into the speaking task when individuals dictate with ASR. The larger the amount of energy that goes into production of the end product (i.e., the dictated message), the lower the efficiency of the work done. Therefore, the high workloads associated with ASR technology decrease the efficiency with which a dictation task is completed.

Reduction of efficiency during dictation with ASR was also evident in the data for duty cycle, inspiratory pause time, and the time required for dictation of a message. The results from the duty cycle analysis suggest that dictation with ASR is not characterized by as efficient use of the expiratory breath supply as during natural dictation. Likewise, inspiratory pause time was significantly longer when dictating with ASR, and these inspirations preceded smaller-than-control expirations for speech, especially for discrete-word ASR. Further, the time it took to generate a message while dictating with ASR was increased, which influenced the overall efficiency with which an end product was generated when using this technology. Finally, efficiency was impaired by the extra time and energy that was spent in dictation of corrective utterances in addition to words representing the content of the speaker's message.

The efficiency issues related to the speech-breathing and speech-production data can be interpreted in context of the kinematic results. It appears that most individuals used primarily the rib cage to effect lung volume exchange and did so within the midvolume range of the lung volume. The kinematic data also indicate that the individuals within this study favored a mechanically advantageous background chest wall configuration while speaking (Hixon, Goldman & Mead, 1973; Hixon, Mead & Goldman, 1976; Hoit, 1995; Hoit et al., 1996) and produced speech in a range that maintained efficient use of muscular capabilities (Hixon, Goldman & Mead, 1973). It also appears that some individuals with SCI may have been more frugal with respiratory muscle energy expenditure as evidenced by the fact that they did not depart far from the relaxation characteristic, which would require muscular effort of the accessory rib cage and neck muscles to assume such a configuration. That they did not recruit the accessory rib cage and neck muscles in such a manner afforded these individuals with SCI more muscular energy for respiratory tasks related to inspiration and 'checking' of relaxation pressures during expiration. All of the aforementioned kinematic observations imply that speakers were efficient in configuring their respiratory system to support speech-breathing and speech-production behaviors during tasks of dictation with ASR. However, the kinematic observations also confirm that lung volume excursions during dictation with discrete-word ASR were small, which impacts other speech-breathing and speech-production variables in terms of efficiency and presumably is a result of the regulating effect of this technology.

Fatigue of the Voice

Of related concern to the investigator in this study was the effect that ASR systems may have on the voice. The increased workload and decreased efficiency associated with dictation using ASR suggest that there is potential for a speaker to be placed at risk for overuse of the respiratory and laryngeal systems and, subsequently, fatigue. Prolonged use of the voice for the working task has been cited as the main loading factor for individuals who use their voices professionally (Vilkman et al., 1999). Corroborating this notion is the fact that changes in the tissue characteristics of the vocal folds following periods of prolonged use of the voice have been documented (Scherer et al., 1987; Mann et al., 1999). Scherer and colleagues (1987) describe the presence of vocal fold edema in two subjects after approximately 1 to 2 hours of prolonged loud phonation. Mann and colleagues (1999) describe conditions such as erythema of the vocal folds, vocal fold edge irregularity, vocal fold edema, and early polypoid swelling associated with a relatively short period (6 days) of vocally demanding work.

The theory surrounding the development of vocal fatigue is related to respiratory and vocal muscle over-exertion resulting from overuse of the voice. Within the current study, overuse of the respiratory system over the course of each dictation task was revealed in the increased number of breath groups that were the result of smaller respiratory pump cycles, slower speaking rates, increased instances of apnea and a decreased duty cycle of the respiratory system during dictation with both ASR systems. Overuse of the voice when dictating with ASR would include speaking for an extended period of time (i.e., more words were needed to express a given message than what would be expected for dictation without ASR). Thus, some of the primary pre-requisite factors for the development of vocal fatigue were evident during dictation with ASR in the present study and suggest that individuals who routinely use this technology may be vulnerable to the development of laryngeal fatigue, and possibly laryngeal pathology. Although the frequency of breath groups and instances of apnea did not appear to differ greatly between dictation with and without ASR when one considers these variables as a function of a per-minute value, the extended length of the dictation task with either ASR system cannot be ignored.

With respect to duration of a speaking task and subsequent overuse of the voice leading to the development of voice disorders, it is useful to interpret the results of this study within the framework of cumulative trauma theory. This theory has been developed in association with research on cumulative trauma disorders of the upper extremities, which are linked with repeated and prolonged muscular activity in combination with abnormal posturing and insufficient rest, resulting in mechanical fatigue due to repeated application of a workload (Goldstein, Armstrong, Chaffin & Matthews, 1987; Kumar, 1999; Kumar, 1994). The results of the current investigation suggest that users of ASR will confront situations in which there will be need for repeated and prolonged muscle activity within the respiratory, laryngeal and articulatory systems for successful speech recognition. If individuals rely on this technology for daily use in an occupational setting, the likelihood is high that the speech mechanism will not be allowed sufficient rest. The National Institute of Occupational Safety and Health (1986) warned that 'when job demands…repeatedly exceed the biomechanical capacity of the worker, the activities become trauma-inducing' (p.19). Thus, it is possible that chronic use of dictation with ASR may provoke a cumulative trauma type of disorder of the vocal mechanism.

Clinical Significance of Outcomes

The importance of this initial investigation of speech breathing and speech production associated with ASR is that it confirms the role of ASR systems in provoking unnatural speaking behaviors and provides a basis for clinical recommendations for users of the technology. Table 6-4 contains information regarding the significant findings across all dependent variables analyzed in this investigation. The clinical picture that emerges from the information collected about speech-breathing and speech-production behaviors is useful in understanding the mechanisms that may lead to vocal disorders in some individuals who use ASR technology.

Dependent Variables	Speaker Group	Dictation Condition	Speech Sample
Mean volume (%VC) per pre-speech inspiration	▨	none = continuous; none > discrete; continuous > discrete	▨
Mean volume (%VC) per expiratory segment for speech	▨	none = continuous; none > discrete; continuous > discrete	▨
Average absolute volume expenditure per expiratory segment for speech (ml)	▨	none = continuous; none > discrete; continuous > discrete	▨
Duty cycle of the respiratory system within the expiratory phase for speech	▨	none > continuous; none > discrete; continuous = discrete	CP>SL
Average duration of inspiratory pause time (ms)	▨	none < continuous; none < discrete; continuous = discrete	▨
Average number of syllables per breath group (California Passage) (#)	▨	none > continuous; none > discrete; continuous > discrete	not tested
Average number of syllables per breath group (Letter) (#)	▨	none > continuous; none > discrete; continuous > discrete	not tested
Frequency of breath groups (California Passage) (#)	▨	none < continuous; none < discrete; continuous < discrete	not tested
Frequency of breath groups (Letter) (#)	▨	none = continuous; none < discrete; continuous < discrete	not tested
Frequency of apnea (California Passage) (#)	▨	none < continuous; none < discrete; continuous < discrete	not tested
Frequency of apnea (Letter) (#)	▨	▨	not tested
Time needed for dictation (California Passage) (minutes)	▨	none < continuous; none < discrete; continuous < discrete	not tested
Number of words spoken (California Passage) (#)	▨	none < continuous; none < discrete; continuous = discrete	not tested
Number of words spoken (Letter) (#)	▨	none > continuous; none > discrete; continuous > discrete	not tested

Table 6- 4. Table of significant differences found between speaker groups, across dictation conditions, and between speech samples. The grey boxes represent hypotheses that were tested but found to be non-significant. The boxes labelled 'not tested' were comparisons that were not included for statistical analysis within the present study. The symbols '<' and '>' indicate the direction of the significant finding across dictation conditions and speech samples for each of the dependent variables. The notation '=' indicates that a significant difference between two of the dictation conditions did not exist. None = no ASR; continuous = continuous-speech ASR; discrete = discrete-word ASR; CP = California Passage; SL = spontaneous letter.

The first important clinical observation from this study relates to vocal fatigue resulting from overuse of the voice when using ASR technology. Based on current literature (Hillman et al., 1989; Koufman & Blalock, 1988; Sander & Ripich, 1983; Sapir, Atias, & Shahar, 1990; Scherer et al., 1987), the primary investigator hypothesized prior to the implementation of this experiment that the factors important in prevention of a voice disorder were avoidance of speaking behaviors characteristic of misuse and overuse of the voice, such as speaking with altered breath support and speaking for extended periods of time. The results in Table 6-4, and those in Tables 6-1 through 6-3, confirm that speech-breathing and speech-production behaviors are altered from normal and that the speaking task is lengthened when using ASR, which provides support for the development of voice disorders associated with the use of this technology. However, the degree to which each system provokes these unnatural behaviors differs. It appears that the advancements of ASR technology that now allow users to dictate more naturally (i.e., with continuous-speech ASR) lead to fewer of these unfavorable speaking behaviors than older ASR technology (i.e., discrete-word ASR). Nonetheless, concerns still exist regarding overuse of the speaking mechanism even with continuous-speech ASR.

The other important clinical observation from this study relates to the performance of the individuals with SCI. Manning and colleagues (1992) studied individuals with and without SCI to determine differences in whole-body oxygen consumption during resistive-loaded breathing tasks. Their results suggested that respiratory muscle oxygen cost was greater in individuals with SCI. These investigators concluded that the decreased efficiency of inspiration in individuals with SCI could compound respiratory muscle weakness and lead to a premature onset of fatigue in situations where the inspiratory muscles are faced with a greater workload. Because results of the current study suggest that ASR is associated with higher speech-production workload demands that require an increased frequency of breathing for speech over the duration of the dictation task, the potential for compounding the load imposed on the respiratory system in individuals with SCI who use ASR must be considered. Notwithstanding the potential for special vulnerability among users with SCI, the results from this study suggest that such participants managed the additional workload in nearly the same fashion as their non-disabled cohorts. Specifically, individuals with SCI were able to generate respiratory volumes that supported a similar amount of speech as their controls, and accomplished the dictation tasks using a similar number of respiratory pump cycles over a similar period of time. Caution is warranted in concluding that no group differences exist, however, as several trends for differences in speaking behaviors were evident between groups and the statistical power associated with the non-significant between-groups findings was low. Therefore, further exploration of speaking behaviors in a greater number of individuals with SCI when using ASR may be required to fully understand the effect of this technology on that population.

The results of the current investigation also provide a basis for clinical recommendations for individuals who use ASR systems. Based on recommendations made for other professional voice users (Hoit et al., 1996), preventative measures that may reduce the potential for development of voice disorders in individuals using ASR can be developed. Such measures include education and training related to vocal hygiene and the proper use of ASR systems. With respect to vocal hygiene, individuals using ASR would benefit from understanding what constitutes vocal misuse/overuse. They could be counseled in the importance of hydration and the risks associated with substances that are known to affect vocal health, including drying agents such as antihistamines and caffeine (Martin, 1984). In addition, it may be useful to include training in normal aspects of respiration, such as beginning utterances at approximately twice the resting tidal breathing depth, ending utterances slightly above REL, inspiring at linguistically-appropriate boundaries, and avoiding breath-holding behavior. Acknowledgment of the possibility for overuse of the voice associated with dictation while using ASR and instruction on the need for intermittent voice rest may be beneficial. Further to this, professional training in the use of ASR systems may be essential for individuals who plan to utilize the technology extensively. Such training may make a user more confident in solving misrecognitions, and may prevent misuse and overuse of the vocal mechanism when speech is not recognized. Confidence in solving misrecognitions has the potential to reduce the cognitive load associated with determining the correct command for correction, which may prevent periods of apnea or air wastage associated with such processing. Confidence in solving misrecognitions will also result in the likelihood that an appropriate correction will be spoken to the computer rather than one that requires further correction. Such confidence may reduce the overall frustration that a user may experience while dictating, which may prevent emotionally induced abusive vocal behaviors such as speaking with an increased pitch or loudness of the voice or beginning phonation by bringing the vocal folds together abruptly. Such behavior would require increased effort from the respiratory and/or laryngeal systems. This increased effort, along with the repetitiveness of the task of dictating with ASR may increase the risk of development of a vocal disorder.

Critique of the Study: Design Validity - Strengths and Limitations

Consideration of the design validity of this research will follow the format that has been put forth by Cook and Campbell (1979) and outlined by Portney and Watkins

(2000). When examining the validity of experimental research, these authors suggest that four areas be considered: internal validity; external validity; statistical conclusion validity; and construct validity of causes and effects. Within the discussion of each of these, the strengths and limitations of this investigation will be explored.

Internal Validity

An experiment is said to be internally valid when there is evidence that the experimental treatment is responsible for the observed changes in the dependent variable. Consideration and control of confounding factors that have the potential to interfere with the relationship between the independent and dependent variables will enhance internal validity. A number of aspects of this investigation were designed to enhance internal validity across speakers, speech samples and dictation conditions. First, potential participants with and without SCI were carefully evaluated to ensure that there were no speech, language, voice or intelligibility problems that could influence the speech-breathing, kinematic, and speech-production outcomes of the study or confound the speech-recognition process. In individuals with SCI, the level of spinal cord lesion was limited to the cervical portion of the cord, and time since injury was controlled so that compensatory changes in respiratory behaviors had stabilized. In addition, the use of a control group helped to rule out extraneous subject effects related to the presence of a spinal cord injury that could influence the findings, especially those related to respiratory volumes, inspiratory pause times, duty cycle, and frequency of apnea. With respect to speech sample, the speaking content during the reading task was controlled by having all participants read the same passage. For the spontaneous sample, subjects were given a specified dictation task (i.e., a letter about their favorite movie) so that content and form could be controlled to a certain degree. With respect to dictation condition, the use of a control speaking condition helped to distinguish speech-breathing and speech-production behaviors that were a function of the introduction of ASR into the speaking environment. Also, random assignment of the order of dictation condition controlled for threats to validity such as history and testing, which will be described shortly. Finally, the research protocol set out prior to the study was designed to control for other extraneous variables such as the time across which data were collected, the location of testing, the positioning of subjects and microphones, and the order of the collection of respiratory data.

Threats to internal validity that are applicable to this investigation include such things as history, maturation, selection and instrumentation. History refers to the effect of specific events that occur after the independent variable has been introduced, or between different testing times. The effect of history in this investigation was minimal because there was limited time (10 minutes) between experimental conditions. However, Portney and Watkins (2000) have suggested that even with short periods of time between measurements, factors such as the subject moving around and the effect of conversation between the subject and experimenter could introduce confounding effects. In the current investigation, subject movement was a concern both within and between dictation conditions because the magnetometers are sensitive to postural changes of the torso and pectoral girdle. Within a specific dictation condition, it is possible that undetected postural shifts could have occurred; these shifts could influence the interpretation of lung volume parameters and background configuration of the chest wall. In addition, subjects with SCI required periods of movements to prevent spasms associated with having to minimize shoulder and torso movements during magnetometric data sampling. If postural shifts were obvious in the data that were being measured for the speech-breathing variables, those data were not included. Reinitializing the magnetometers just before a subject began to dictate in a new condition controlled for subject movement between conditions. Regardless, it is possible that small shifts during dictation went undetected and this may have a bearing on volume-related results. With respect to conversation, the primary investigator physically left the recording booth during the rest time between dictation conditions so that subjects would rest their voices and would not be tempted to discuss the previous dictation condition or other aspects of the experiment. Another issue related to history within the current study involved the effect that a preceding dictation condition may have had on the following one. Randomizing the order of dictation conditions for each subject and allowing a rest period of ten minutes between each dictation condition were used to control for any carry-over effects.

The second issue related to threats to internal validity in the present investigation is that of maturation. Maturation occurs simply as a function of the passage of time and may cause changes in subjects' responses on a repeated measurement. Maturation is typically an issue when a larger amount of time has passed between experimental measurements than occurred in this investigation. However, because the experimental recording session was a rather lengthy one (2.5 to 3 hours), maturation becomes an issue in that boredom or fatigue of the subjects with the passage of time could have affected the results. Randomization of the order of dictation conditions across subjects controlled for one condition's being influenced by such effects more than the others.

Issues relating to selection arise when random assignment to groups is not used, as was the case in this study. The subjects selected for inclusion in the study were volunteers and included both males and females within each group. If subjects volunteered for the study because they were particularly interested in ASR or computers, the results of the study may be biased as such individuals may perform better than others who may not have the same interest level in such things. For example, individuals who are more expert users of computers may be more comfortable in problem-solving interactions with them (Molnar & Kletke, 1996). During dictation with ASR, this comfort may result in less frustration and therefore, an ability to more efficiently command the computer to correct misrecognitions (Molnar & Kletke, 1996) resulting in a reduction of the number of words spoken, the number of breath groups produced, and the amount of time spent dictating.

The inclusion of both sexes has implications for the accuracy of speech recognition and the respiratory results. With respect to ASR recognition, differences in both the fundamental frequency of the voice and the resulting distribution of harmonics in the radiated speech spectrum could plausibly create differences in the accuracy with which male versus female speech is recognized by the computer. Although scarce, research has suggested that men are better recognized by ASR systems than women (Noyes & Frankish, 1989), while another study suggests there is no difference (Brown & Vosburgh, 1989). With respect to speech-breathing, research has shown that men and women differ in volumes of air expired during speech, with men expiring a greater volume per syllable than women (Hoit et al., 1989). In the current study, men and women were included in one group because of the limited sample that was attainable. For measures of absolute volume expenditure within each group, the inclusion of both male and female data has the potential to result in grand mean values that will be lower than what might be expected for men and higher than what might be expected for women. When average absolute volume expenditure was calculated for women and men across dictation conditions and speech samples within the current study, it was observed that women expended less volume than men (367 ml and 435 ml, respectively). This would not be an issue between groups, however, as potential differences were addressed with control subjects that were matched for sex and height. Further to this, because each subject served as his/her own control across dictation conditions and speech samples, the inclusion of both males and females was not felt to be a detriment for the collection of the speech-breathing and speech-production data beyond the issue raised for absolute volume expenditure within groups. And finally, volumes were adjusted to %VC as a form of normalization for the sex differences. Regardless, future studies may want to include sex of the subjects as an independent variable.

Lastly, issues related to the instrumentation used in this study have the potential to influence the results. Digital analysis of speech breathing traces is likely to be less error prone than the visual analog method used in the current investigation. Although measurement of the speech-breathing traces that were displayed on the oscilloscope screen was done with great care and measurement reliability was high, it is possible that visual errors on the part of the primary investigator could result in under- or overestimation of actual volume displacements.

External Validity

External validity is related to the extent to which results can be generalized to individuals, settings, and times that are different from those that characterized this experiment. A feature that strengthens the external validity of this study is that two of the most popular speech recognition software packages on the market at the time of the study, one representing discrete-word technology and the other representing continuous-speech technology, were used in collection of the dictation data. In addition, the discrete-word and continuous-speech packages were made by the same manufacturer, which helps to control for the confounding effects that may be related to different software platforms (i.e., different search engines, modeling techniques, grammars and vocabularies), which have the potential to affect ease of use and accuracy of speech recognition. With respect to the participants in the study, the subjects were limited to a specific age range (25-55 years) so that generalization of results could be made to other individuals in that age range without the confounding influence of age-related phenomena on some of the measures that were made. The speech and breathing characteristics of individuals much older or younger than the subjects in this study are reported to differ with respect to respiratory volumes, initiation and termination levels in the lung volume, syllables produced per breath group (Hoit & Hixon, 1987; Hoit, Hixon, Watson & Morgan, 1990; Sperry & Klich, 1992), and acoustic characteristics of the voice (Lee, Potamianos & Narayanan, 1999; Ramig & Ringel, 1983).

Threats to external validity that are applicable to this investigation include such things as the interaction of treatment with selection and the interaction of treatment with history. The interaction of treatment with selection refers to the potential for characteristics of the subjects who were selected for the study to interfere with generalization to a wider population. For example, the individuals with SCI had a vested interest in working with the technology as most had indicated that they were considering it for use at work or home, whereas individuals in the control group were more likely to indicate that they were just curious about the technology. With a more vested interest in the task, individuals with SCI may have made a greater effort to perform the task with attentiveness and diligence, which may result in a more positive interaction with the software and possibly more efficient dictation. In addition, two of the individuals with SCI had used ASR systems prior to participation in the study, although their use of such systems was limited. Because it was plausible that potential participants with SCI would have had some exposure already to ASR systems, a decision was made to include individuals who had been exposed to the technology, but to select for those with "limited exposure". However, even minimal familiarity with the systems may have influenced the results for these subjects. Finally, the subjects who participated did not receive any formal training outside of the session with the primary investigator in the use of ASR technology. Professional training has the potential to increase the ease with which individuals use a system (J. Vargo, personal communication, January 19, 2001). Thus, the results from the participants in this study may not be generalizable to individuals who have received professional training or have extensive experience with ASR software, with or without training.

Another threat to external validity within the current study is the interaction of treatment with history. This issue becomes important when considering whether the results of this experiment can be generalized to a different period of time, past or future. The results of the current study can be generalized only to individuals using the ASR software packages employed herein, but not necessarily to other packages that were previously or are currently on the market. Further to this, the findings are also not generalizable to ASR packages of the future that may incorporate statistical modeling techniques superior to the ones currently available. Such advancements have the potential to increase recognition accuracy and allow for the production of even more natural speech when interacting with such systems. ASR technology that allows speakers to dictate in a manner that rivals conversational speech likely will reduce the unnatural speaking behaviors that were documented in this study.

Statistical Conclusion Validity

Statistical conclusion validity is concerned with the accuracy with which results can be interpreted based on the statistical tests used to analyze the data. A feature that strengthens the statistical conclusion validity of the current investigation is that most variables were normally-distributed and met the homogeneity of variance assumptions for analyses with ANOVA. Although some variables did not meet the necessary homogeneity of variance assumptions, samples were of equal size and therefore, the analysis of variance statistic was considered to be powerful enough to account for any reasonable departures from the assumptions of normality and homogeneity so that the validity of inferences drawn from the data were not seriously affected (Portney and Watkins, 2000). Further, for any data that did not meet assumptions of sphericity, a Greenhouse-Geisser adjustment was made, making it harder to achieve significance and thereby, compensating for the bias towards a Type I error that may be associated with unequal variances in repeated-measures ANOVAs (Portney & Watkins, 2000; SPSS Applications Guide, 1999). In addition, in order to minimize the familywise error rate of the multiple ANOVA comparisons within each portion of this investigation (i.e. the probability that at least one comparison, in a collection of comparisons, will include a Type I error), a Bonferroni correction was applied within each set of data. It could be argued that the Bonferroni corrections should have been applied to the data as a whole (i.e., all 10 dependent variables) rather than separately to each set of dependent variables (i.e., 5 speech-breathing and 5 speech-production variables). Applying a Bonferroni correction to all 10 variables would have made it harder to find significance; however, follow-up analyses revealed that all significant findings within this study would have remained significant even with a Bonferroni correction for 10 comparisons. Thus, the significant findings within this study could have withstood stricter correction procedures without affecting the results. In future investigations, reducing the number of dependent variables will increase the power of the results. Within the current investigation, the retrospective power associated with the comparisons made across dictation conditions was high, which provides confidence that important clinical differences have been detected (Ottenbacher & Maas, 1999).

Although statistical power was high for some analyses, it was low for others, which may lead to a situation where significant effects are not realized because of inadequate sample size. Within the current study, the between-group analyses were characterized by low retrospective power. However, the use of retrospective power to reach conclusions about study findings has been criticized in that this type of power is a function of the error variance and effect size of the statistical analysis and thereby only restates the degree of significance of the study findings (i.e., highly significant findings will have high retrospective power and highly non-significant findings will have low power) (Thomas, 1997). The prospective power analysis completed within this study included a power level set at .70 for a large effect size. It has been suggested that an 80% power level represents reasonable protection against Type II error (Portney & Watkins, 2000). Hence, the prospective power level may have been set somewhat too low and the interpretation of the study results, especially those related to the between-group analyses, should be completed with caution.

Where trends in the data were detected in the direction of the expected differences stated in the hypotheses, low statistical power could have been responsible for the lack of significant between-group differences that were expected, based on existing literature. For example, it was surprising that the measures related to volume did not differ between groups as such differences have been reported by others (Hoit et al, 1990). Also, differences in inspiratory time between individuals with SCI and their non-disabled cohorts have been described in the literature (Hixon & Putnam, 1983) but were not a significant between-group finding in this research. These discrepancies suggest that statistical conclusion validity is questionable for the between-group comparisons herein and that future investigations with more subjects are required to confirm or refute these findings.

Construct Validity of Causes and Effects

Construct validity is concerned with the theoretical conceptualization of the dependent and independent variables. With respect to the current study, the theoretical concepts related to speech-breathing behaviors have a long history and are well established in the literature. Thus, the techniques used to acquire the data and the interpretation of those data are based on sound principles, which strengthens the construct validity of this experiment.

Threats to construct validity are related to how variables within an experiment are operationally defined and to potential biases introduced into a study by the participants or the experimenter (Portney and Watkins, 2000). Although the breathing data collected within this investigation were operationally defined, the validity of the conversion from rib cage and abdomen diameters to volumes of air exchanged relies on the assumption that the rib cage and abdomen move with a single degree of freedom and that no other degrees of freedom are introduced (Hoit et al., 1990). Undetected postural shifts made by the subjects may have violated this assumption, and therefore volume estimates may have been under- or overestimated. Also, the validity of the kinematic data for participants with SCI must be questioned as results related to background configuration of the chest wall and patterns of chest wall part contribution to lung volume exchange do not agree with other reports for representatives of this same population of individuals (Hoit et al., 1990). Abdominal magnetometer placement on individuals with SCI is not as straightforward as it is for individuals whose torsos have not been distorted by the effects of SCI on musculo-skeletal alignments. It is possible that improper placement of the abdominal magnetometer could lead to transduction of what really is lower rib cage diameter change. This would violate the assumption that each magnetometer was measuring displacement of a separate chest wall part. Finally, the validity of the concept of respiratory fatigue during speech must be addressed. Although respiratory fatigue is an entity that has been alluded to in association with laryngeal fatigue (Titze, 1984), there is no support in the literature to substantiate fatigue of the respiratory system during speech. This could be addressed in the future by studying PImax and PEmax values collected before and after strenuous speaking tasks.

With respect to bias, the participants in the study may have been subject to experimental bias that results in a Hawthorne effect; that is, individuals who were singled out for inclusion could perform better because of the expectations created by the situation. With respect to the experimenter, experimental bias may lead her to unintentionally give clues as to what the expected results should be. An attempt was made by the primary investigator to reduce the possibility for giving participants clues as to what her *a priori* hypotheses were by remaining neutral when discussing the different aspects of the systems and how to work with them. During the experimental session, conversation with a participant regarding how easy or difficult dictation was with one system or another was avoided.

Finally, construct validity also is concerned with the researcher's goals and how well the results of measurement based on such goals can be generalized to the population represented by participants in the study. Within the present investigation, it was the goal of the researcher to take a 'snapshot' of the speech and breathing behaviors of fairly naïve users of ASR in one experimental session. Thus, generalization is limited because the performances of these subjects were observed on only one occasion after they had received minimal training, making it difficult to generalize the outcomes from this study to the speech and breathing events that might occur with professional training or over a longer period of time spent learning and using ASR software. With professional training, user confidence may be improved resulting in dictation that is more efficient (e.g., fewer words spoken as commands for correction and shortened dictation times).

Conclusions

The findings of this investigation reveal that when ASR becomes part of the speech servomechanism, speech-breathing and speech-production behaviors are regulated in a manner that differs from natural speech. This regulation results in increased workloads for the speech-production system, which has the potential to result in laryngeal fatigue, and may in part be responsible for anecdotal reports of vocal difficulties in individuals using ASR. To this point however, the vocal difficulties linked with ASR use have been associated only with the use of discrete-word ASR in non-disabled individuals (Kambeyanda et al., 1997). It was the intent of this study to understand more about this association, not only in relation to discrete-word ASR, but also in relation to newer technology (i.e., continuous-speech ASR) that boasts the ability to 'understand' more naturally produced speech. In addition, it was important to examine the impact of both of these technologies not only on non-disabled individuals, but also on individuals whose speech-production systems have been affected by injury (i.e., SCI).

With respect to the two types of ASR technologies, evidence of an increased workload was more pronounced during dictation with discrete-word ASR, and therefore the potential for the development of a voice disorder is likely greater with this type of technology. Continuous-speech ASR has made gains over discrete-word ASR in more closely approximating a natural speaking situation. Nevertheless, evidence of a greater-than-normal speaking workload was apparent when individuals dictated with continuous-speech ASR and thus, concern regarding vocal fatigue is still warranted for users of this type of ASR. This concern is warranted for both non-disabled individuals and those with SCI, with slightly greater concern for individuals with SCI. Although the results of this study did not reveal a difference between these two groups of individuals, the statistical power of the between-group analyses was low and therefore, true group differences may not have been detected. In addition, trends in the speech-breathing and speech-production data for individuals with SCI suggested that workload was higher for them than for individuals in the control group, though not significantly so. Further, because the respiratory system of individuals with SCI may already be predisposed to fatigability (Hopman et al., 1997; Sinderby et al., 1996), an increased workload on this system is likely to compound the problem of fatigue. If the speaking apparatus of both non-disabled individuals and those with SCI continues to be stressed under conditions of increased workload, and subsequently fatigue, cumulative trauma of the vocal fold tissues may occur and may result in the development of soft-tissue lesions such as vocal fold nodules (Hillman et al., 1989; Mann et al., 1999). The occurrence of vocal fatigue or a laryngeal pathology such as nodules may result in dysphonias that render the voice useless for dictation purposes and leave the user with a significant communication disability.

From a human factors perspective, the addition of ASR into the speech servomechanism changes speech-breathing and speech-production behaviors of both non-disabled individuals and those with SCI such that efficiency of these behaviors within the task is compromised. This regulating effect of ASR has the potential to prolong speaking time for the user, to increase energy expenditure during the task, and to eventually result in fatigue. Continued research efforts to develop more sophisticated statistical pattern recognition techniques, as well as to continue to model the intricacies of connected speech so that articulation, rate and prosody are more natural when dictating, should serve to increase the efficiency with which an individual can dictate a message using ASR. Theoretically, this should result in a decrease in speech-breathing and speech-production behaviors that impose a heavier-than-normal workload on the speaking apparatus.

Due to the potential for development of vocal disorders with ASR use, the issues of increased workload and decreased efficiency should be addressed by both the individuals who use the technology and by those who manufacturer it. If these issues are not addressed, serious vocal consequences that have the potential to impact vocal health, and thus productivity of users of ASR in the workplace, may result. Users of the technology need to be cognizant of the amount of time they spend dictating, the amount of rest they allow their voice, and what characterizes good vocal hygiene. Manufacturers of the technology need to continue to develop research initiatives to improve ASR software so that speakers can dictate using speech that more closely approximates natural production. If both parties assume these responsibilities, the potential for the development of voice disorders when using ASR should be minimized.

Although this research has provided a plausible basis for vocal fatigue that may result from altered speaking behaviors during dictation with ASR, there is still much to understand about the human-computer interface in order to appreciate fully the impact that ASR systems will have on the voice and other aspects of the speech system. Following are some suggestions for future research that will enhance our knowledge of the effect of ASR systems on the speaking apparatus.

Suggestions for Future Study

This study has provided a framework upon which future studies of the effects of ASR on the user can be based. There are many aspects of the user-computer interaction that are yet to be explored. With respect to speech-breathing, more specific information related to lung volume initiation and termination levels during dictation would be useful, as abnormally high or low initiation levels have been related to vocal problems. As well, knowledge related to breath holding could be furthered by understanding where individuals hold their breath in the speech-production process – is it random? does it occur after misrecognitions? or does it occur when an individual is formulating a command to the ASR system? With respect to speech production, useful information would include the analysis of segmental durations of speech while dictating with ASR so that issues related to articulatory changes could be understood more comprehensively. For example, it would be interesting to document changes in durations of vowels and consonants when repeating a misunderstood word, and to determine what type of articulatory changes lead to recognition. Finally, to pursue the issue of vocal muscle misuse, future studies of ASR could assess evidence of misuse that one may see clinically (Morrison, 1997) in users of ASR. Assessment could include observation and measurement of:

- Rigid posture of head and neck
- Jaw tension related with anxiety or other emotional factors
- Tense and asymmetrical shoulder position
- Head extension with jaw displaced forward
- Restricted jaw motion
- Scalloped tongue edges
- Decreased vertical laryngeal movement
- Tense strap muscles and suprahyoid tension
- Increased subglottic air pressure
- Increased perception of vocal effort as reported by the patient

Over and above the aforementioned areas of interest is the issue of professional training in the use of ASR and the effect that such instruction may have on the speaking patterns observed during dictation. Also, it will be important to understand if and how these speech-breathing and speech-production behaviors change over time.

Finally, acoustic and perceptual analyses will yield valuable information regarding changes in the speech signal across dictation conditions. Acoustic analyses will provide insight into how ASR systems affect the loudness, pitch, and voice-onset rise time of the voiced signal. In addition, inverse filtering could be applied to the acoustical signal associated with dictation in different conditions to understand more about glottal source characteristics such as the speed of closure of the vocal folds and sharpness of glottal closure (i.e., whether the vocal folds contact one another in an instantaneous way or in a more gradual fashion along their length and depth) (Chasaide & Gobl, 1997). Perceptual analyses will provide insight into the naturalness of speech during dictation with ASR as perceived by listeners, and results could be referenced to acoustic analyses. It also would be interesting to gain some insight into the user's perception of tension experienced during dictation with ASR systems. In concert with this, it would be interesting to monitor physiological stress such as the blood pressure and heart rate of individuals speaking under different dictation conditions to determine if there is a physiologic basis for the frustration that is occasionally reported during use of these systems, and if alterations in these physiologic data correspond to changes in the acoustic signal of the voice.

Further exploration of the speech-breathing and speech-production behaviors associated with ASR use, as well as investigation of related acoustical, perceptual and ergonomic issues will enhance understanding of the regulating effects of ASR. In addition, the role that ASR may play in the development of voice disorders will be better understood through such knowledge.

References

Almenoff, P.L., Spungem, A.M., Lesser, M. & Bauman, W.A. (1995). Pulmonary function survey in spinal cord injury: Influences of smoking and level and completeness of injury. Lung, 173, 297-306.

Andrews, M. & Shank, K.H. (1983). Some observations concerning the cheering behavior of school-girl cheerleaders. Language, Speech, and Hearing Services in Schools, 14, 150-156.

Arnaut, G. (1995, September 15). Talking to computers has its hazards. The Globe and Mail, C4.

Astridge, P.S., Kaye, G.C., Whitworth, S., Kelly, P., Camm, A.J., & Perrins, E.J. (1993). The response of implanted dual chamber pacemakers to 50 Hz extraneous electrical interference, Pacing and Clinical Electrophysiology, 16, 1966-74.

Attanasio, J.S. (1987). Relationships between oral sensory feedback skills and adaptation to delayed auditory feedback. Journal of Communication Disorders, 20, 391-402.

Baber, C., Mellor, B., Graham, R., Noyes, J.M. & Tunley, C. (1996). Workload and the use of automatic speech recognition: the effects of time and resource demands. Speech Communication, 20, 37-53.

Bland, M. (1995). An Introduction to Medical Statistics. Oxford, NY: Oxford University Press.

Boone, D.R. & McFarlane, S.C. (1988). The Voice and Voice Therapy. Englewood Cliffs, NJ: Prentice-Hall.

Borden, G.J. & Harris, K.S. (1984). Speech Science Primer: Physiology, Acoustics, and Perception of Speech. Baltimore, MD: Williams & Wilkins.

Bourke, M.E. (1996). The patient with a pacemaker or related device. Canadian Journal of Anaesthesia, 43, R24-41.

Brobeck, J.R. (1965). Exchange, control, and regulation. In J.R. Brobeck & W.S. Yamamoto (Eds.), Physiological controls and regulations. Philadelphia, PA: WB Saunders.

Brown, N.R. & Vosburgh, A.M. (1989). Evaluating the accuracy of a large-vocabulary speech recognition system. Proceedings of the Human Factors Society 33rd annual Meeting, 296-300.

Buchler, B., Magder, S. & Roussos, C. (1985). Effects of contraction frequency and duty cycle on diaphragmatic blood flow. Journal of Applied Physiology, 58, 265-273.

Chan, R.W. & Titze, I.R. (2000). Viscoelastic properties of human vocal fold mucosa: theoretical characterization based on constitutive modeling. Journal of the Acoustical Society of America, 107, 565-580.

Chasaide, A. & Gobl, C. (1997). Voice source variation. In W.J. Hardcastle & J. Laver (Eds.), The Handbook of Phonetic Sciences. Cambridge, MA: Blackwell.

Cook, T.D. & Campbell, D.T. (1979). Quasi-experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton Mifflin.

Crystal, T.H. & House, A.S. (1988a). Segmental durations in connected-speech signals: current results. Journal of the Acoustical Society of America, 83, 1553-1573.

Crystal, T.H. & House, A.S. (1988b). Segmental durations in connected-speech signals: syllabic stress. Journal of the Acoustical Society of America, 83, 1574-1585.

Dalston, R., Warren, D. & Smith, L. (1990). The aerodynamic characteristics of speech produced by normal speakers and cleft palate speakers with adequate velopharyngeal function. Cleft Palate Journal, 27, 393-401.

Davis, A.W. (1996). Safety & productivity improvements with voice terminals. American Society of Safety Engineers, March, 37-39.

Dragon Systems Inc. (2000). [On-line]. Available: www.naturallyspeaking.com

Dubno, J.R., Ahlstrom, J.B. & Horwitz, A.R. (2000). Use of context by young and aged adults with normal hearing. Journal of the Acoustical Society of America,107, 538-546.

Estenne, M. & De Troyer, A. (1986). The effects of tetraplegia on chest wall statics. American Review of Respiratory Disease, 134, 121-124.

Eustace, C.S., Stemple, J.C. & Lee, L. (1996). Objective measures of voice production in patients complaining of laryngeal fatigue. Journal of Voice, 10, 146-154.

Fairbanks, G. (1955). Selective vocal effects of delayed auditory feedback. Journal of Speech and Hearing Disorders, 20, 333-346.

Fleiss, J.L. (1986). The Design and Analysis of Clinical Experiments. New York, NY: John Wiley & Sons.

Flynn, M.C. & Dowell, R.C. (1999). Speech perception in a communicative context: an investigation using question/answer pairs. Journal of Speech, Language, and Hearing Research, 42, 540-552.

Forner, L. & Hixon, T. (1977). Respiratory kinematics in profoundly hearing-impaired speakers. Journal of Speech and Hearing Research, 20, 373-408.

Fosler-Lussier, E. & Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in convertional speech. Speech Communication, 29, 137-158.

Fucci, D., Crary, M., Warren, J.A. & Bond, Z. (1977). Interaction between auditory and oral sensory feedback in speech regulation. Perceptual and Motor Skills, 45, 123-129.

Fujiwara, T., Hara, Y. & Chino, N. (1999). Expiratory function in complete tetraplegics; Study of spirometry, maximal expiratory pressure, and muscle activity of pectoralis major and latissimus dorsi muscles. American Journal of Physical Medicine and Rehabilitation, 78, 464-469.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. Journal of the Acoustical Society of America, 63, 223-230.

Gelfer, M.P., Andrews, M.L., & Schmidt, C.P. (1991). Effects of prolonged loud reading on selected measures of vocal function in trained and untrained singers. Journal of Voice, 5, 158-167.

Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in spontaneous speech. New York, NY: Academic Press.

Goldstein, S., Armstrong, T., Chaffin, D. & Matthews, L. (1987). Analysis of cumulative strain in tendons and tendon sheaths. Journal of Biomechanics, 20, 1-6.

Gooderham, M. (1995, September 14). High-tech RSI aid creates new problem. The Globe and Mail, A12.

Gotaas, C. & Starr, C.D. (1993). Vocal fatigue among teachers. Folia Phoniatrica, 45, 120-129.

Greenberg, S. (1999). Speaking in shorthand - a syllable-centric perspective for understanding pronunciation variation. Speech Communication, 29, 159-176.

Helfer, K.S. (1997). Auditory and auditory-visual perception of clear and conversational speech. Journal of Speech, Language, and Hearing Research, 40, 432-443.

Herrington-Hall, B.L., Lee, L., Stemple, J.C., Niemi, K.R., & McHone, M.M. (1988). Description of laryngeal pathologies by age, sex, and occupation in a treatment-seeking sample. Journal of Speech and Hearing Research, 53, 57-64.

Higgins, M.B., Carney, A.E., & Schulte, L. (1994). Physiological assessment of speech and voice production of adults with hearing loss. Journal of Speech and Hearing Research, 37, 510-521.

Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M. & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. Journal of Speech and Hearing Research, 32, 373-392.

Hixon, T. (1987). Respiratory function in speech. Respiratory Function in Speech and Song. Boston, MA: College-Hill Press. pp. 1-54.

Hixon, T., Goldman, M. & Mead, J. (1973). Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung. Journal of Speech and Hearing Research, 16, 78-115.

Hixon, T. & Hoit, J. (1984). Differential subsystem impairment, differential motor system impairment, and decomposition of respiratory movement in ataxic dysarthria: a spurious trilogy. Journal of Speech and Hearing Disorders, 49, 435-441.

Hixon, T. & Hoit, J. (1985). Reply to Abbs (1985). Journal of Speech and Hearing Disorders, 50, 312-317.

Hixon, T., Mead, J. & Goldman, M. (1976). Dynamics of the chest wall during speech production: function of the thorax, rib cage, diaphragm, and abdomen. Journal of Speech and Hearing Research, 19, 297-356.

Hixon, T. & Putnam, A.H.B. (1983). Voice disorders in relation to respiratory kinematics. Seminars in Speech and Language, 4, 217-231.

Hixon, T., Putnam, A., & Sharp, J. (1983). Speech production with flaccid paralysis of the rib cage, diaphragm, and abdomen. Journal of Speech and Hearing Disorders, 48, 315-327.

Hixon, T. & Weismer, G. (1995). Perspectives on the Edinburgh study of speech breathing. Journal of Speech and Hearing Research, 38, 42-60.

Hodge, M.M. & Rochet, A.P. (1989). Characteristics of speech breathing in young women. Journal of Speech and Hearing Research, 32, 466-480.

Hoit, J.D. (1995). Influence of body position on breathing and its implications for the evaluation and treatment of speech and voice disorders. Journal of Voice, 9, 341-347.

Hoit, J.D., Banzett, R.B., Brown, R. & Loring, S.H. (1990). Speech breathing in individuals with cervical spinal cord injury. Journal of Speech and Hearing Research, 33, 798-807.

Hoit, J. & Hixon, T. (1987). Age and speech breathing. Journal of Speech and Hearing Research, 30, 351-366.

Hoit, J. Hixon, T., Altman, M. & Morgan, W. (1989). Speech breathing in women. Journal of Speech and Hearing Research, 32, 353-365.

Hoit, J. Hixon, T., Watson, P. & Morgan, W. (1990). Speech breathing in children and adolescents. Journal of Speech and Hearing Research, 33, 51-69.

Hoit, J.D., Jenks, C.L., Watson, P.J. & Cleveland, T.F. (1996). Respiratory function during speaking and singing in professional country singers. Journal of Voice, 10, 39-49.

Hood, L.J. (1998). An overview of neural function and feedback control in human communication. Journal of Communication Disorders, 31, 461-470.

Hopman, M., van der Woude, L., Dallmeijer, A., Snoek, G. & Folgering, H. (1997). Respiratory muscle strength and endurance in individuals with tetraplegia. Spinal Cord, 35, 104-108.

Horii, Y. & Cooke, P. (1978). Some airflow, volume, and duration characteristics of oral reading. Journal of Speech and Hearing Research, 21, 470-481.

Jiang, J.J., Lin, E. & Hanson, D.G. (2000). Vocal fold physiology. Otolaryngologic Clinics of North America, 33, 699-718.

Jobe, F.W. (1976). Screening vision in schools. Newark, DE: International Reading Association.

Kambeyanda, D., Singer, L., & Cronk, S. (1997). Potential problems associated with use of speech recognition products. Assistive Technology, 9, 95-101.

Kent, R.D. & Read, C. (1992). The Acoustic Analysis of Speech. San Diego, CA: Singular.

Kitch, J.A. & Oates, J. (1994). The perceptual features of vocal fatigue as self-reported by a group of actors and singers. Journal of Voice, 8, 207-214.

Kitch, J.A., Oates, J. & Greenwood, K. (1996). Performance effects on the voices of 10 choral tenors: acoustic and perceptual findings. Journal of Voice, 10, 217-227.

Konno, K. & Mead, J. (1967). Measurement of the separate volume changes of the rib cage and abdomen during breathing. Journal of Applied Physiology, 22, 407-422.

Koufman, J.A. & Blalock, P.D. (1988). Vocal fatigue and dysphonia in the professional voice user: Bogart-Bacall syndrome. Laryngoscope, 98, 493-498.

Kroemer, K.H. & Grandjean, E. (1997). Fitting the Task to the Human: A Textbook of Occupational Ergonomics. Bristol, PA: Taylor & Francis.

Kumar, S. (1994). A conceptual model of overexertion, safety, and risk of injury in occupational settings. Human Factors, 36, 197-209.

Kumar, S. (1999). Selected theories of musculoskeletal injury causation. In S. Kumar (Ed.). Biomechanics in Ergonomics. Philadelphia, PA: Taylor & Francis.

Laine, T., Warren, D., Dalston, R., Hairfield, W. & Morr, K. (1988). Intraoral pressure, nasal pressure and airflow rate in cleft palate speech. Journal of Speech and Hearing Research, 31, 432-437.

Larson, C. R. (1998). Cross-modality influences in speech motor control: the use of pitch shifting for the study of F0 control. Journal of Communication Disorders, 31, 489-503.

Laver, J. (1980). The Phonetic Description of Voice Quality. New York, NY: Cambridge University Press.

Lee, S., Potamianos, A. & Narayanan, S. (1999). Acoustics of children's speech: developmental changes of temporal and spectral patterns. Journal of the Acoustical Society of America, 105, 1455-1468.

Leeper, H. (1976). Voice initiation characteristics of normal children and children with vocal nodules: a preliminary investigation. Journal of Communication Disorders, 9, 83-94.

Leinonen, L., Hiltunen, T., Linnankoski, I. & Laakso, M. (1997). Expression of emotional-motivational connotations with a one-word utterance. Journal of the Acoustical Society of America, 102, 1853-1863.

Lippmann, R.P. (1997). Speech recognition by machines and humans. Speech Communication, 22, 1-15.

Loveridge, B., Sanii, R. & Dubo, H. (1992). Breathing pattern adjustments during the first year following cervical spinal cord injury. Paraplegia, 30, 479-488.

Ludlow, C.L. & Cikoja, D. (1998). Is there a self-monitoring speech perception system?. Journal of Communication Disorders, 31, 505-510.

Makhoul, J. & Schwartz, R. (1995). State of the art in continuous speech recognition. Proceedings of the National Academy of Science USA, 92, 9956-9963.

Manning, H., McCool, F., Scharf, S., Garshick, E. & Brown, R. (1992). Oxygen cost of resistive-loaded breathing in quadriplegia. Journal of Applied Physiology, 73, 825-31.

Martin, F.G. (1984). Drugs and the voice. In V.L. Lawrence (Ed.), Transcript of the Twelfth Symposium: Care of the Professional Voice, Part I: Scientific Papers. New York, NY: The Voice Foundation.

McCool, F.D., McCann, D.R., Leith, D.E. & Hoppin, F.G. (1986). Pressure-flow effects on endurance of inspiratory muscles. Journal of Applied Physiology, 60, 299-303.

Milenkovic, P. (1992). CSpeech: Computer speech waveform acquisition editing and analysis. Madison, WI: University of Wisconsin.

Miller, M. & Verdolini, K. (1995). Frequency and risk factors for voice problems in teachers of singing and control subjects. Journal of Voice, 9, 348-362.

Mitchell, H.L., Hoit, J.D. & Watson, P.J. (1996). Cognitive-linguistic demands and speech breathing. Journal of Speech and Hearing Research, 39, 93-104.

Molnar, K.K. & Kletke, M.G. (1996). The impacts on user performance and satisfaction of a voice-based front-end interface for a standard software tool. International Journal of Human-Computer Studies, 45, 287-303.

Moon, J.B. & Folkins, J.W. (1991). The effects of auditory feedback on the regulation of intraoral air pressure during speech. Journal of the Acoustical Society of America, 90, 2992-2999.

Moon, J.B., Folkins, J.W., Smith, A.E., & Luschei, E.S. (1993). Air pressure regulation during speech production. Journal of the Acoustical Society of America, 94, 54-63.

Moore, R.K. (1999). Automatic speech recognition in the marketplace. [On-line]. Available: www.iee.org.uk.

Morrison, M. (1997). Pattern recognition in muscle misuse voice disorders: how I do it. Journal of Voice, 11, 108-114.

Murphy, C.H. & Doyle, P.C. (1987). The effects of cigarette smoking on voice-fundamental frequency. Otolaryngology-Head and Neck Surgery, 97, 376-380.

National Institute for Occupational Safety and Health (1986). Proposed national strategies for the prevention of leading work-related diseases and injuries: Part 1. Washington, DC: Association of Schools of Public Health, 19.

Neils, L.R. & Yairi, E. (1987). Effects of speaking in noise on vocal fatigue and vocal recovery. Folia Phoniatrica, 39, 104-112.

Noyes, J.M. & Frankish, C.R. (1989). Gender differences in speech recognition performance. Proceedings of European Conference on Speech Communication and Technology, 1, 110-112.

Ottenbacher, K. & Maas, F. (1999). How to detect effects: statistical power and evidence-based practice in occupational therapy research. American Journal of Occupational Therapy, 53, 181-188.

Padilla, E. (1998). January 1998 Speech Recognition Update. 21st Century Eloquence. Available Website: http://voicerecognition.com/1998/trends/

Percy, A.K., Zoghbi, H.Y., Lewis, K.R. & Jankovic, J. (1988). Rett syndrome: qualitative and quantitative differentiation from autism. Journal of Child Neurology, 3 Suppl, S65-67.

Picheny, M.A., Durlach, N.I. & Braida, L.D. (1989). Speaking clearly for the hard of hearing III: an attempt to determine the contribution of speaking rate differences in intelligibility between clear and conversational speech. Journal of Speech and Hearing Research, 32, 600-603.

Portney, L.G. & Watkins, M.P. (2000). Foundations of Clinical Research. Connecticut: Appleton & Lange.

Putnam, A.H.B. & Hixon, T.J. (1983). Respiratory kinematics in speakers with motor neuron disease. In M. McNeil, J. Rosenbek & A. Aronson (Eds.), The Dysarthrias. San Diego, CA: College-Hill Press.

Putnam, A.H.B., Shelton, R.L., & Kastner, C.U. (1986). Intraoral air pressure and oral air flow under different bleed and bite-block conditions. Journal of Speech and Hearing Research, 29, 37-49.

Ramig, L.A. & Ringel, R.L. (1983). Effects of physiological aging on selected acoustic characteristics of the voice. Journal of Speech and Hearing Research, 26, 22-30.

Reich, A. McHenry, M., & Keaton, A. (1986). A survey of dysphonic episodes in high-school cheerleaders. Language, Speech, and Hearing Services in Schools, 17, 63-71.

Rowan, G. (1997, December 1). Microsoft aims for speech recall software. The Globe and Mail, B3.

Russell, B.A., Cerny, F.J. & Stathopoulos, E.T. (1998). Effects of varied vocal intensity on ventilation and energy expenditure in women and men. Journal of Speech, Language, and Hearing Research, 41, 239-248.

Sander, E.K. & Ripich, D.E. (1983). Vocal fatigue. Annals of Otology, Rhinology, and Laryngology, 92, 141-145.

Sapir, S. (1993). Vocal attrition in voice students: survey findings. Journal of Voice, 7, 69-74.

Sapir, S., Atias, J., & Shahar, A. (1990). Symptoms of vocal attrition in women army instructors and new recruits: Results from a survey. Laryngoscope, 100, 991-994.

Saraclar, M., Nock, H. & Khudanpur, S. (2000). Pronunciation modeling by sharing Gaussian densities across phonetic models. Speech Communication, 14, 137-160.

Scherer, R.C., Titze, I.R., Raphael, B.N., Wood, R.P., Ramig, L.A., & Blager, R.F. (1987). Vocal fatigue in a trained and an untrained voice user. In T. Baer, C. Sasaki, & K. Harris (Eds.), Laryngeal function in phonation and respiration (pp.533-555). Boston, MA: College-Hill Press.

Sheard, C., Adams, R. & Davis, P. (1991). Reliability and agreement of ratings of ataxic dysarthric speech samples with varying intelligibility. Journal of Speech and Hearing Research, 34, 285-293.

Shrout, P. & Fleiss, J. (1979). Intraclass correlation: uses of assessing rater reliability. Psychological Bulletin, 86, 420-428.

Sinderby, C., Weinberg, J., Sullivan, L., Lindstrom, L. & Grassino, A. (1996). Electromyographical evidence for exercise-induced diaphragm fatigue in patients with chronic cervical cord injury or prior poliomyelitis infection. Spinal Cord, 34, 594-601.

Sitaram, R.N. & Screenivas, T. (1997). Incorporating phonetic properties in hidden Markov models for speech recognition. Journal of the Acoustical Society of America, 102, 1149-1158.

Sivak, E., Shefner, J. & Sexton, J. (1999). Neuromuscular disease and hypoventilation. Current Opinion in Pulmonary Medicine, 5, 355-362.

Sluijter, A.M. & Van Heuven, V.J. (1996). Spectral balance as an acoustical correlate of linguistic stress. Journal of the Acoustical Society of America, 100, 2471-2485.

Smith, E., Gray, S.D., Dove, H., Kirchner, L. & Heras, H. (1997). Frequency and effects of teachers' voice problems. Journal of Voice, 11, 81-87.

Solomon, N.P. & Hixon, T. (1993). Speech breathing in Parkinson's disease. Journal of Speech and Hearing Research, 36, 294-310.

Sorenson, D. & Horii, Y. (1982). Cigarette smoking and voice fundamental frequency. Journal of Communication Disorders, 15, 135-144.

Sperry, E.E. & Klich, R.J. (1992). Speech breathing in senescent and younger women during oral reading. Journal of Speech and Hearing Research, 35, 1246-1255.

Stathopoulos, E.T. & Sapienza, C. (1993). Respiratory and laryngeal function of women and men during vocal intensity variation. Journal of Speech and Hearing Research, 36, 64-75.

Stemple, J.C., Stanley, J. & Lee, L. (1995). Objective measures of voice production in normal subjects following prolonged voice use. Journal of Voice, 9, 127-133.

Strik, H. & Cucchiarini, C. (1999). Modeling pronunciation variation for ASR: a survey of the literature. Speech Communication, 29, 225-246.

Tausz, A. (1997a, October 21). Voice dictation software gives you the last word. The Globe and Mail, C4.

Tausz, A. (1997b, October 21). Worth shouting about? The Globe and Mail, C5.

Thomas, L. (1997). Retrospective power analysis. Conservation Biology, 11, 276-280.

Titze, I.R. (1984). Vocal fatigue: Some biomechanical considerations. In V.L. Lawrence (Ed.), Transcript of the Twelfth Symposium: Care of the Professional Voice, Part I: Scientific Papers. New York, NY: The Voice Foundation.

Titze, I.R. (1993). Current topics in voice production mechanisms. NCVS Status & Progress Report, 4, 303-310.

Titze, I.R. (1994). Mechanical stress in phonation. Journal of Voice, 8, 99-105.

Titze, I.R. & Sundberg, J. (1992). Vocal intensity in speakers and singers. NCVS Status Report and Progress Report, 2, 17-37.

Uchanski, R.M., Choi, S.S., Braida, L.D., Reed, C.M. & Durlach, N.I. (1996). Speaking clearly for the hard of hearing IV: further studies of the role of speaking rate. Journal of Speech and Hearing Research, 39, 494-509.

van Kuijk, D. & Boves, L. (1999). Acoutic characteristics of lexical stress in continuoustelephone speech. Speech Communication, 27, 95-111.

Vilkman, E. (2000). Voice problems at work: a challenge for occupational safety and health arrangement. Folia Phoniatrica et Logopaedica, 52, 120-125.

Warren, D.W. (1986). Compensatory speech behaviors in individuals with cleft palate: a regulation/control phenomenon? Cleft Palate Journal, 23, 251-260.

Warren, D.W. (1988). Aerodynamics of Speech. In NJ Lass (ed.). Handbook of Speech-Language Pathology and Audiology. Toronto, Ontario: B.C. Decker Inc.

Warren, D.W., Allen, G., & King, H.A. (1984). Physiologic and perceptual effects of induced anterior open bite. Folia Phoniatrica, 36, 164.

Warren, D. W., Dalston, R., Morr, K., Hairfield, W. & Smith, L. (1989). The speech regulating system: temporal and aerodynamic responses to velopharyngeal inadequacy. Journal of Speech and Hearing Research, 32, 566-575.

Warren, D.W., Hall, D.J., & Davis, J. (1981). Oral port constriction and pressure-airflow relationships during sibilant productions. Folia Phoniatrica, 33, 380-394.

Warren, D.W., Morr, K.E., Rochet, A.P. & Dalston, R.M. (1989). Respiratory system response to a decrease in velopharyngeal resistance. Journal of the Acoustical Society of America, 86, 917-924.

Warren, D.W., Rochet, A.P., Dalston, R.M., & Mayo, R. (1992). Controlling changes in vocal tract resistance. Journal of the Acoustical Society of America, 91, 2947-2953.

Weinstein, C.J. (1995). Military and government applications of human-machine communication by voice. Proceedings of the National Academy of Science USA, 92, 10011-10016.

Wilson, F.B. (1990). Voice Profile. Bellingham, WA: Voice Tapes, Inc.

Winkworth, A.L. & Davis, P.J. (1997). Speech breathing and the Lombard Effect. Journal of Speech, Language, and Hearing Research, 40, 159-169.

Winkworth, A.L., Davis, P. J., Adams, R.D. & Ellis, E. (1995). Breathing patterns during spontaneous speech. Journal of Speech and Hearing Research, 38, 124-144.

Winkworth, A.L., Davis, P. J., Ellis, E. & Adams, R.D. (1994). Variability and consistency in speech breathing during reading: lung volumes, speech intensity, and linguistic factors. Journal of Speech and Hearing Research, 37, 535-556

APPENDIX A

Information For Potential Participants

(Flesch-Kincaid Grade Level: 8.0)

Title of Project: Speaking behaviors associated with the use of automatic speech recognition systems by individuals with and without spinal cord injury.

Investigators: Jana Rieger, University of Alberta, 492-7588
Anne Rochet, University of Alberta, 492-9674

Purpose: The purpose of this study is to look at breathing and speaking behaviors that people use when dictating into two different types of automatic speech recognition (ASR) systems. The data are being collected for a graduate thesis.

Background: ASR technology allows people to use speech to interact with a computer. Instead of using a keyboard and mouse, ASR makes it possible to do word-processing and other computer-related tasks by talking into a microphone. ASR software is relatively new. Little is known about the speaking behaviors that occur when people use it. There is some concern that improper use of ASR for a long period of time may affect the voice. Thus, we need to record speech and breathing information while using ASR from men and women who are 20 to 55 years old. Two groups of people will participate: those with and without spinal cord injury. This will help us understand the speaking behaviors that people use with ASR and how they relate to the health of the voice.

Procedures: Participation involves 3 visits to Corbett Hall on the University of Alberta campus.

- On visit 1, you will fill out a brief form about your health and language. Your hearing and vision will be tested. If you meet the study requirements, you will then be asked to train the computer to recognize your speech by reading a list of words and a short story. You will wear a simple headset microphone while you read. Visit 1 will last no more than 90 minutes.

- During visit 2, you will be able to see how the computer recognizes your speech. You will also learn how to use the ASR programs for word-processing tasks. Visit 2 will last no more than 90 minutes.

- During visit 3, you will use ASR for word-processing. You will be asked to read some short readings and to describe a picture. Your breathing and speech will be recorded on a tape recorder while you read. This last visit is the longest. It will take about 2 hours.

To measure your breathing, you will be asked to breathe in and out a few times on an instrument (respirometer) that measures the amount of air you inhale and exhale. Then, you will have 4 small discs (called magnetometers, about the size of quarters) attached to your skin with double-sided tape. One disc will be attached in front to the middle of your chest and another just above your belly button. The other two discs will be attached to the middle of your back at the same level as the discs on the front. If you are male, it may be necessary for the experimenter to shave a small patch of hair on your chest so that the discs can stick to your skin. The hair will grow back.

To be included in this study, you must speak and read English fluently. You must have good hearing and good vision.

Certain conditions are known to affect speech and breathing and may prevent you from participating. Therefore, you may not participate if you have a neurological disorder such as a stroke or Parkinson's disease or a structural disorder such as cleft palate. If any speech problems are discovered that exclude you from participating, a referral to a speech pathologist can be made if you so wish. You may not participate if you are a smoker. You may not participate if you have a pacemaker or any other electronic implant.

If you are eligible to participate, you may not do so while you have a cold or laryngitis.

If you are a potential participant with spinal cord injury, we can only accept you for the study if you are quadriplegic but are able to breathe on your own. Many individuals with quadriplegia have a lot of muscle spasms but can predict a few seconds before hand that a spasm is coming on. If this description fits you, we ask that you let us know if you feel a spasm coming on. This is because the spasm will be picked up by the magnetometers. This will distort the results. This is important because we want to measure your breathing patterns and not your muscle spasms.

If any further analysis is conducted with the study, further ethics approval will be sought first.

Benefits: You will provide information about the effects of ASR technology on certain aspects of speech. This information may be useful in preventing voice disorders among people who use ASR. We acknowledge that 5 hours of your time is needed to participate. We hope that the chance to use ASR during that time will be valuable to you.

Risks: There are no known risks involved. You will be asked to sit as still as possible for certain periods of time (but will be allowed to take 'wiggle' breaks).

Reimbursement: All participants will receive a monetary token to offset travel and parking costs for each visit. Payment will be provided to you immediately following the completion of each session. If you decide to withdraw and the session is not completed, a percentage of the payment will be given to you based upon the time spent in the session. You will not receive payment for unattended sessions.

Confidentiality: The investigator and her supervisor will be the only people with access to the records of this study. They will store them in a secure place for a period of seven years after the study is completed and erase them according to University regulations. They will report the results in written form and in lectures. You will never be identified in any report or presentation. All information will be held confidential except when professional codes of ethics and or legislation require reporting.

Freedom to Withdraw: Your participation is voluntary. You may withdraw at any time without consequence.

Additional Contacts: If you have any concerns about any aspect of this study, you may contact Dr. Sharon Warren at 492-7856.

APPENDIX B

Subject's History Form: Language, Speech And Voice

(Flesch-Kincaid Grade Level: 3.7)

Name: ______________________________
Date: ____________________
Birthdate: (D/M/Y) _____/_____/_____
Birthplace: __________________________
Address:
____
____

1. What is your native language (mother tongue):____________________
____

2. Briefly describe your occupation (duties, hours of work, etc.): __________
____
____

3. Describe your daily voice use and the importance of your voice in these situations:
Work:____
Home:____
Social/Recreation:______________________________

4. Have you ever used automatic speech recognition software?____
If yes, which program have you used?__________________________

Do you use the ASR program (choose one):

Less than 5 hours per week______________
5 hours per week______________________
More than 5 hours per week______________

If you use it more than 5 hours per week, please estimate the number of hours you would typically use it during one week ____________________

5. If you do use automatic speech recognition software, would you consider yourself:

A beginning user, with limited skills____________________
An intermediate user, with average skills________________
An expert user, with above average skills________________

6. Have you ever been diagnosed with a voice disorder?____ If yes, please name the disorder, and the approximate date of diagnosis: _______________
____

7. Do you experience heartburn?____ If yes, how often?:__________________
State cause of heartburn if known ________________________________

8. Do you experience any of the following?:
- Hoarse/husky voice yes____ no____ If yes, how often?_______
- Breathy voice yes____ no____ If yes, how often?_______
- Loss of voice yes____ no____ If yes, how often?_______
- Pitch breaks yes____ no____ If yes, how often?_______
- Reduced loudness yes____ no____ If yes, how often?_______
- Unsteady voice yes____ no____ If yes, how often?_______
- Tension in your neck yes____ no____ If yes, how often?_______
- Throat/neck pain yes____ no____ If yes, how often?_______
- Throat fatigue yes____ no____ If yes, how often?_______
- Throat tightness yes____ no____ If yes, how often?_______
- Throat pain when talking
yes____ no____ If yes, how often?_______
- Throat pain when swallowing
yes____ no____ If yes, how often?_______
- Increased need to cough/clear throat
yes____ no____ If yes, how often?_______
- Running out of breath while talking
yes____ no____ If yes, how often?_______

9. Have you ever lost your voice?____ If yes, what happened?:
____
____

10. Indicate any other voice problems you experience and how often they occur:____
____

11. Have you ever received any professional help from a Speech-Language Pathologist?____ If yes, please explain:__________________________
____

12. Have you ever received any professional speaking or singing instruction?___
If yes, describe (approximate date, activities): ______________________
____

13. Do you have any allergies? _____ Please list:______________________
____

14. Do you presently have a cold or other respiratory illness? ____ If yes, describe:____

15. Have you ever been diagnosed with a neurological condition (e.g. Parkinson's Disease, Multiple Sclerosis)?____ If yes, describe:______________________

____

16. Are you currently taking any medications?____ If yes, please list: __________

____

____

____

17. Do you smoke?____ If yes, how much? ______________________________

18. If you have a history of any of the following disorders, please provide details.

Speech Disorder:

____

Hearing Disorder:

____

Breathing Disorder:

____

Visual Disorder:

____

Cleft of Lip and/or Palate:

____

19. Have you sustained a spinal cord injury?____ If yes, please answer the following:

What year did the accident occur? __________________________

At what level of your spine did the injury occur (e.g. C3, L1)? _____________

Is the lesion complete?________ incomplete?_________ don't know_______

Please describe the use of your hands (e.g. limited use, some use of right/left hand):____

____

APPENDIX C
Consent Template

Part 1:

Title of Project: Speaking behaviours associated with the use of automatic speech recognition systems by individuals with and without spinal cord injury

Principal Investigator(s): Jana Rieger, University of Alberta, 492-7588

Co-Investigator(s): Anne Rochet, University of Alberta, 492-9674

Part 2:

Do you understand that you have been asked to be in a research study?	Yes	No
Have you read and received a copy of the attached Information Sheet?	Yes	No
Do you understand the benefits and risks involved in taking part in this research study?	Yes	No
Have you had an opportunity to ask questions and discuss this study?	Yes	No
Do you understand that you are free to refuse to participate or withdraw from the study at any time? You do not have to give a reason and it will not affect your access to ASR training elsewhere or to other assistive technologies.	Yes	No
Has the issue of confidentiality been explained to you? Do you understand who will have access to your records?	Yes	No
Do you want the investigator(s) to inform your family doctor, physiotherapist, or caregiver that you are participating in this research study?	Yes	No

If so, please provide their name:______________________________

This study was explained to me by: ______________________________

I agree to take part in this study.

______________________________ _______________ ______________________________

Signature of Research Participant   Date   Witness

______________________________ ______________________________

Printed Name   Printed Name

I believe that the person signing this form understands what is involved in the study and voluntarily agrees to participate.

______________________________ _______________

Signature of Investigator or Designee   Date

THE INFORMATION SHEET MUST BE ATTACHED TO THIS CONSENT FORM AND A COPY GIVEN TO THE RESEARCH SUBJECT

APPENDIX D
California Passage

California is a unique state. It is one of the few states that has all the geographical features found in the rest of the country including deserts, forests, mountain ranges, and beaches. Its beaches draw thousands and thousands of people each year particularly during the summer months when the sun is shining, the skies are blue, and the ocean is warm enough to swim in. Surfers are often in the water by daybreak. Of course there are many other things to do besides surfing such as sailing, swimming, water skiing, kite flying, and sun bathing. In the winter the mountains of California are favorite vacation spots. Here, snow skiing is the sport. There are many places in California to snow ski but the largest and most popular is Mammoth Mountain. Because of its popularity the property surrounding the Mammoth ski resort is extremely expensive. Unfortunately, the threat of earthquakes in this area is very high. In fact, earthquakes are common occurrences in many parts of California. Because of this, there are people who are afraid that someday a large piece of the state will fall into the Pacific Ocean. The possibility of a serious earthquake such as the one that demolished San Francisco in 1906 frightens some people enough that they choose not to visit California just for that reason.

APPENDIX E

Dictating A Letter

Dictate a letter to your friend and tell them about your favorite movie. Here are some things that you might want to include:

Who were the characters?

When did it take place?

Where did it take place?

What was the story about?

What problems did the characters encounter?

How were those problems resolved?

What was your overall impression of the movie?

Would you recommend it? Why? Why not?